NASA Contractor Report 188242

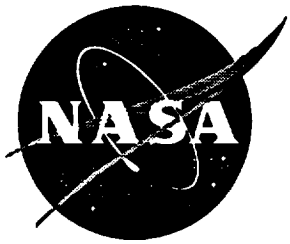
National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program—1992

Volume 2

Richard B. Bannerot and Stanley H. Goldstein, Editors

Grant NGT 44-005-803

December 1992



(NASA-CR-188242-Vol-2) NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA)/AMERICAN SOCIETY FOR ENGINEERING EDUCATION (ASEE) SUMMER FACULTY FELLOWSHIP PROGRAM, 1992, VOLUME 2 (NASA) 165 p

N93-26070
--THRU--
N93-26083
Unclas

G3/99 0157420

NASA Contractor Report 188242

National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program—1992

Volume 2

Richard B. Bannerot, Editor
Mechanical Engineering Department
University of Houston--University Park
Houston, Texas

Stanley H. Goldstein, Editor
University Programs Office
Lyndon B. Johnson Space Center
Houston, Texas

Grant NGT 44-005-803

December 1992

National Aeronautics and Space Administration
Lyndon B. Johnson Space Center
Houston, Texas

Preface

The 1992 Johnson Space Center (JSC) National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program was conducted by the University of Houston and JSC. The program at JSC, as well as the programs at other NASA Centers, was funded by the Office of University Affairs, NASA Headquarters, Washington, D.C. The objectives of the program, which began nationally in 1964 and at JSC in 1965, are

1. To further the professional knowledge of qualified engineering and science faculty members

2. To stimulate an exchange of ideas between participants and NASA

3. To enrich and refresh the research and teaching activities of participants' institutions

4. To contribute to the research objectives of the NASA Centers

Each faculty fellow spent at least 10 weeks at JSC engaged in a research project in collaboration with a NASA/JSC colleague. This document is a compilation of the final reports on the research projects done by the faculty fellows during the summer of 1992. Volume 1 contains reports 1 through 12, and Volume 2 contains reports 13 through 24.

PRECEDING PAGE BLANK NOT FILMED

CONTENTS

Volume 1

Volume 2

PRECEDING PAGE BLANK NOT FILMED

PRECEDING PAGE BLANK NOT FILMED

N93-26071

A Vision Architecture for the Extravehicular Activity Retriever

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1992

Johnson Space Center

Prepared by:	Michael Magee, Ph.D.
Academic Rank:	Professor
University & Department:	University of Wyoming Computer Science Department Laramie, Wyoming 82071-3682
NASA/JSC	
Directorate:	Engineering
Division:	Automation and Robotics
Branch:	Intelligent Systems
JSC Colleague:	Thomas W. Pendleton
Date Submitted:	August 7, 1992
Contract Number:	NGT-44-005-803

ABSTRACT

The Extravehicular Activity Retriever (EVAR) is a robotic device currently being developed by the Automation and Robotics Division at the NASA Johnson Space Center to support activities in the neighborhood of the Space Shuttle or Space Station Freedom. As the name implies, the Retriever's primary function will be to provide the capability to retrieve tools, equipment or other objects which have become detached from the spacecraft, but it will also be able to rescue a crew member who may have become inadvertently de-tethered. Later goals will include cooperative operations between a crew member and the Retriever such as fetching a tool that is required for servicing or maintenance operations.

This report documents a preliminary design for a Vision System Planner (VSP) for the EVAR that is capable of achieving visual objectives provided to it by a high level task planner. Typical commands which the task planner might issue to the VSP relate to object recognition, object location determination, and obstacle detection. Upon receiving a command from the task planner, the VSP then plans a sequence of actions to achieve the specified objective using a model-based reasoning approach. This sequence may involve choosing an appropriate sensor, selecting an algorithm to process the data, reorienting the sensor, adjusting the effective resolution of the image using lens zooming capability, and/or requesting the task planner to reposition the EVAR to obtain a different view of the object.

An initial version of the Vision System Planner which realizes the above capabilities using simulated images has been implemented and tested. The remaining sections describe the architecture and capabilities of the VSP and its relationship to the high level task planner. In addition, typical plans that are generated to achieve visual goals for various scenarios will be discussed. Specific topics to be addressed will include object search strategies, repositioning of the EVAR to improve the quality of information obtained from the sensors, complementary usage of the sensors and redundant capabilities.

INTRODUCTION

Vision systems that provide autonomous or semi-autonomous robots with information that describes their surrounding environment or objects

in that environment should be able to plan and execute actions that solve visual problems efficiently and effectively. From a software architectural design standpoint, the highest level or supervisory planner is called the Task Planner (Figure 1). The Task Planner oversees the actions of several subplanners, one of which is the Vision System Planner. Each of these subplanners can be considered to be an expert with special knowledge regarding how to solve problems within its particular domain. When commanded to do so by the Task Planner, a subplanner will determine a method for achieving the specified goal given its knowledge of the current state of the world and it will then communicate the result of executing the planned action back to the Task Planner.

Although each subplanner is subservient to the Task Planner, it may nevertheless ask for assistance from the Task Planner if such assistance would help it achieve the specified goal. For example, if the Task Planner requests the Vision System Planner to recognize an object and the robot on which the vision hardware is mounted is poorly positioned to sense the object, the VSP may request the Task Planner to cause the robot to be moved. If the Task Planner honors the request from the VSP, it would then send commands to other subplanners (involving navigation and control) to move the robot so that the Vision System can accomplish the objective originally requested by the Task Planner.

The Vision System module itself (Figure 2) should be a self-contained entity capable of accomplishing many types of objectives such as object detection, recognition, tracking and pose estimation. A typical plan that would be formulated to achieve one of these goals would involve choosing an appropriate sensor, selecting an algorithm to process the data, and communicating the results or a request for assistance to the Task Planner. The remaining sections discuss a suggested architecture for such a Vision System within the context of the Extravehicular Activity Retriever.

VISION SYSTEM PLANNER DESIGN CONSIDERATIONS

The planning mechanisms developed are founded on the assumption that there should be at least two visual sensors which provide intensity (color) and range images. There are several reasons why such a multisensory approach is desirable, three of which are particularly significant. First, the availability of sensors with complementary capabilities permits the VSP to select a sensor/algorithm combination that is most appropriate for

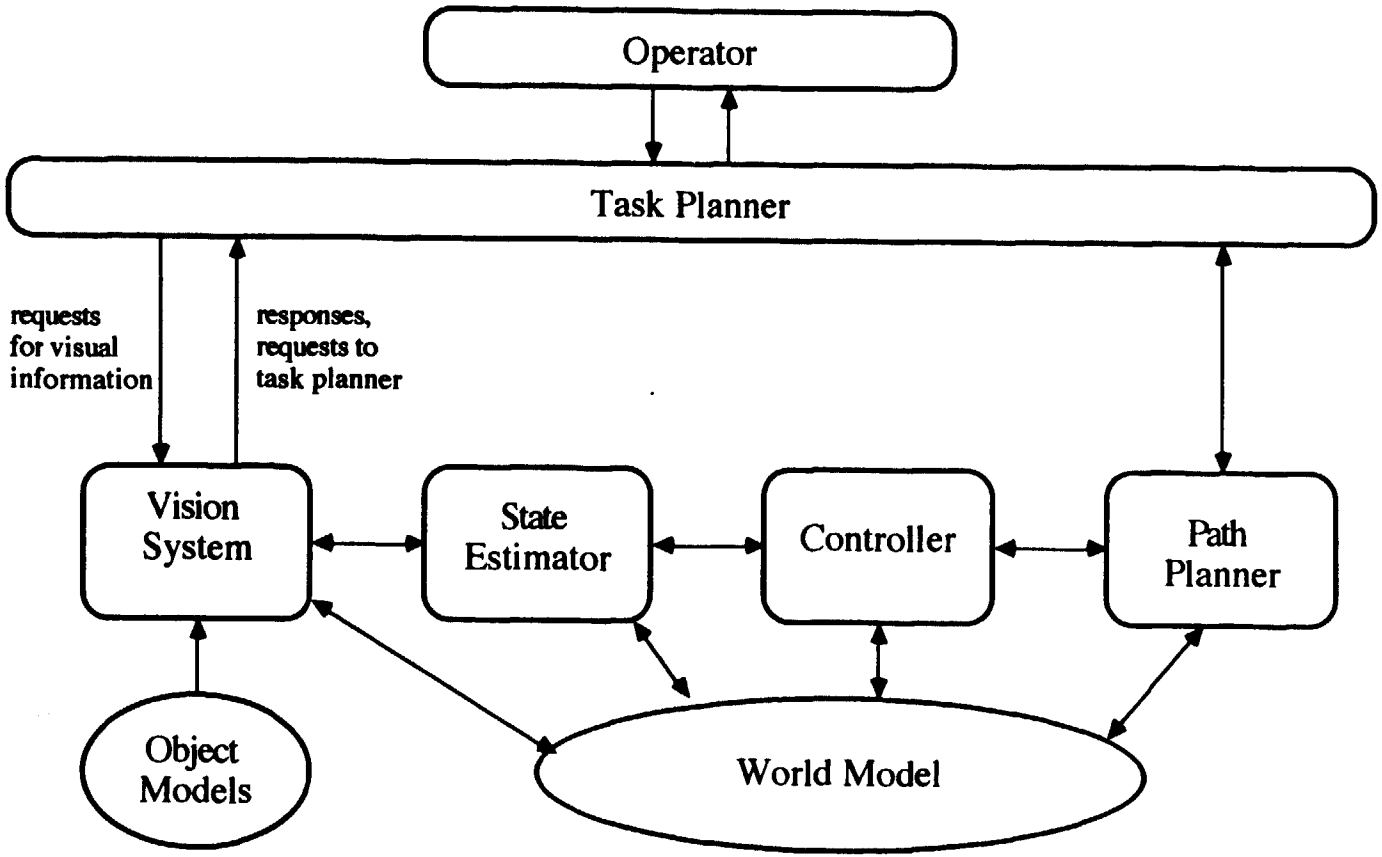


Figure 1: Planning System Architecture

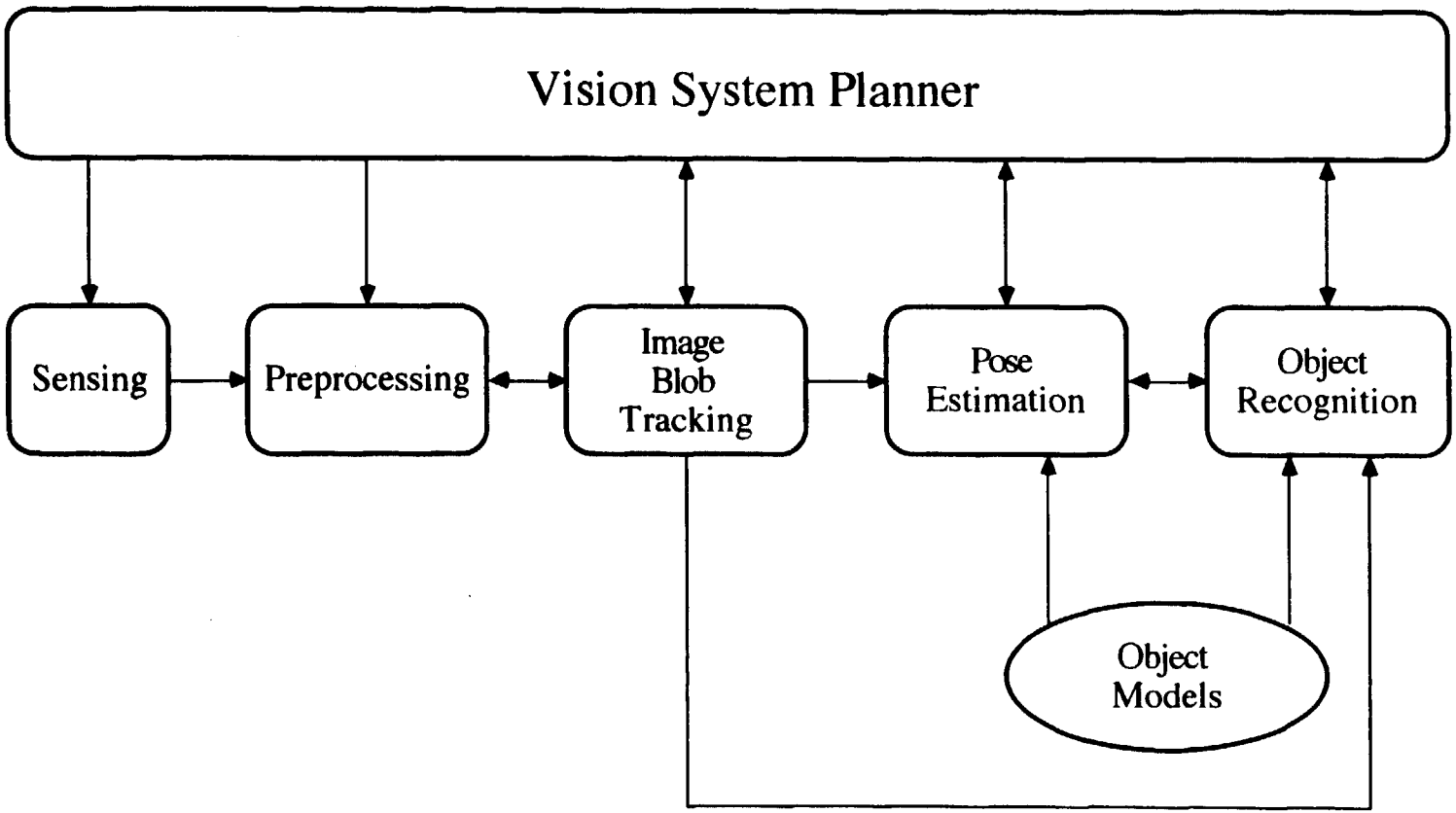


Figure 2: Vision System Architecture

achieving the current visual goal as specified by the task planner. Second, if the sensor that the VSP would normally select as its first choice to achieve the goal is either unavailable or inappropriate for usage because of some current constraint, it may be possible to perform the desired task using the other sensor to achieve the same goal, albeit perhaps by accepting a penalty in performance. Finally, instances may occur for which it is desirable to verify results from two different sensory sources.

The first of the above motivations addresses achieving the visual goal in the most effective manner by allowing the VSP to choose among sensors with complementary capabilities. For example, if it is desired to distinguish between two objects of similar structure with the color of the objects being the primary differentiating feature, then it is apparent that the color camera should be used as the primary sensor. On the other hand, if the size and/or geometry of the objects are most useful for determining identity, then it is important to be able to expeditiously extract and process three-dimensional coordinates. Clearly, this is a task that would be most appropriately assigned to the laser scanner.

The previous example involving the need for three-dimensional coordinates is illustrative of a case in which the primary sensor (the laser scanner) is engaged to extract the required information. However, there may be cases for which the laser scanner cannot be used to obtain range information because (a) the object to be processed is covered with a highly specularly reflective material thus preventing acquisition of good return signals, (b) the laser scanner is currently assigned to another task, or (c) the laser scanner is temporarily not functioning properly. For such instances, it is highly desirable to provide a redundant capability by using the other sensor if possible. The classical method for determining three-dimensional coordinates from intensity images involves a dual (stereo vision) camera setup in which feature correspondences are established and the stereo equations are solved for each pair of feature points. Although the current simulated configuration has only one intensity image camera, this alternative mechanism for computing range values is in fact possible for the VSP to achieve by requesting the task planner to reposition the EVAR such that the camera's initial and final positions are offset by a known baseline distance. Of course, there is a penalty in performance if the (pseudo) stereo vision method is chosen, since the EVAR must be moved and feature correspondences computed. However, it is nevertheless important to have such a redundant sensing capability for the reasons previously mentioned and to be able to independently verify the results

obtained from one sensor or to increase the confidence of those results.

With respect to increasing the confidence of computed results, a typical scenario might involve a case in which the EVAR is close enough to a target object to hypothesize its class based on color, but too far away to definitively recognize its geometric structure using laser scanner data. In this case, the VSP would tentatively identify the object (using color) and would advise the task planner to move closer to the object so that a laser scanner image with higher resolution can be obtained. The confidence of the initial hypothesis would then be strengthened (or perhaps weakened) depending on the conclusion reached by processing the range data at close proximity.

The fundamental architecture for the Vision System includes modules which are designed to detect, recognize, track, and estimate the pose of objects. Upon receiving a request from the main task planner to achieve one of these objectives, the Vision System Planner determines an appropriate sequence of goals and subgoals that, when executed, will accomplish the objective. The plan generated by the VSP will generally involve (a) choosing an appropriate sensor, (b) selecting an efficient and effective algorithm to process the image data, (c) communicating the nominal (expected) results to the task planner or informing the task planner of anomalous (unexpected) conditions or results, and (d) advising the task planner of actions that would assist the vision system in achieving its objectives. The specific plan generated by the VSP will primarily depend on knowledge relating to the sensor models (e.g. effective range of operation, image acquisition rate), the object models (e.g. size, reflectivity, color), and the world model (e.g. expected distance to and attitude of objects). The next section presents the resulting plans generated by the VSP for several different scenarios.

RESULTS

The operation of the VSP that was designed and implemented can best be understood by examining the plans that it generates for various scenarios.

Scenario 1:

State of the world:
Three objects are somewhere in front of the EVAR. One of them is an Orbital Replacement Unit (ORU) with a known color.

Command received by the VSP:
Search in front of the EVAR for an ORU.

Plan generated by the VSP:

1. Search the hemisphere in front of the EVAR by activating the color camera, fixing the effective focal length and spiralling outward from the center until the object is found.

2. If the ORU is found, terminate the (spiralling) search and iteratively refine the estimate of where the object is located by adjusting the sensor gimbals toward the object and reduce the field of view (telephoto zoom) until the object is centered and large in the image.

   If the ORU was not found, the VSP reports failure, in which there are several actions that could be taken. First, the forward hemisphere could be rescanned at higher magnification (a slower process since more scans will be required). Second, the forward hemisphere could be rescanned with increased illumination (requiring a decision to be made regarding the desirability in terms of overall objectives and power consumption by the illumination source). Finally, the VSP could request the Task Planner to rotate the EVAR by 180 degrees and scan the rear hemisphere.

Scenario 2:

State of the world:
Same as Scenario 1

Command received by the VSP:
Determine the distance to the ORU, no sensor specified.

Plan generated by the VSP:

1. Locate the ORU as in Scenario 1 using the color camera.

2. Examine the object model for an ORU and determine which sensor is the most appropriate to be used. In this case, since an ORU is not specularly reflective, the laser scanner is chosen.

3. Examine that part of the laser scanner image that corresponds to the region belonging to the ORU in the color image and compute the distance to those range image elements.

Scenario 3:

State of the world:
Same as Scenario 1

Command received by the VSP:
Determine the distance to the ORU, but force the estimation of distance using single camera lateral stereo vision.

Plan generated by the VSP:

1. Locate the ORU as in Scenario 1 using the color camera.

2. Move the EVAR left a known distance, take an image, and record the location of the ORU in that image. Then move the EVAR right a known distance, take an image, and record the location of the ORU in that image.

4. Using triangulation (stereo vision with two cameras separated by a known baseline distance) compute the distance to the ORU.

Scenario 4:

State of the world:
Same as Scenario 1

Command received by the VSP:
Determine the distance to the ORU and move toward the ORU along the optical axis of the color camera until the EVAR is a specified distance (D) away from it.

Plan generated by the VSP:

1. Locate the ORU as in Scenario 1 using the color camera.

2. Estimate the distance to the ORU (D_{oru}) using the laser scanner.

3. Compute a vector along the optical axis of the color camera whose length is (D_{oru} - D). Transform that vector into EVAR coordinates and move to that position, maintaining the same attitude.

Scenario 5:

State of the world:
Same as Scenario 1

Command received by the VSP:
As in Scenario 4, determine the distance to the ORU and check to determine whether any other objects in the field of view are closer to the EVAR than the ORU prior to moving toward it.

Plan generated by the VSP:

1. Locate the ORU as in Scenario 1 using the color camera.

2. Estimate the distance to the ORU using the laser scanner.

3. Search the range image for values that lie outside of the region containing the ORU and report a potential obstacle if any of the values indicate distances between the EVAR and the ORU.

N93-26072

A STUDY OF MAPPING EXOGENOUS KNOWLEDGE REPRESENTATIONS INTO CONFIG

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1992

Johnson Space Center

Prepared by:	Blayne E. Mayfield
Academic Rank:	Assistant Professor
University and Department:	Oklahoma State University Computer Science Department Stillwater, Oklahoma 74078-0599
NASA/JSC	
Directorate:	Engineering
Division:	Automation and Robotics
Branch:	Intelligent Systems
JSC Colleague:	Jane T. Malin
Date Submitted:	August 4, 1992
Contract Number:	NGT-44-005-803

ABSTRACT

Qualitative reasoning is reasoning with a small set of qualitative values that is an abstraction of a larger and perhaps infinite set of quantitative values. The use of qualitative and quantitative reasoning together holds great promise for performance improvement in applications that suffer from large and/or imprecise knowledge domains. Included among these applications are the modeling, simulation, analysis, and fault diagnosis of physical systems.

Several research groups continue to discover and experiment with new qualitative representations and reasoning techniques. However, due to the diversity of these techniques, it is difficult for the programs produced to exchange system models easily. The availability of mappings to transform knowledge from the form used by one of these programs to that used by another would open the doors for comparative analysis of these programs in areas such as completeness, correctness, and performance.

A group at the Johnson Space Center (JSC) is working to develop CONFIG, a prototype qualitative modeling, simulation, and analysis tool for fault diagnosis applications in the U.S. space program. The availability of knowledge mappings from the programs produced by other research groups to CONFIG may provide savings in CONFIG's development costs and time, and may improve CONFIG's performance. The study of such mappings is the purpose of the research described in this paper.

Two other research groups that have worked with the JSC group in the past are the Northwest University Group and the University of Texas at Austin group. The former has produced a qualitative reasoning tool named SIMGEN, and the latter has produced one named QSIM. Another program produced by the Austin group is CC, a preprocessor that permits users to develop input for eventual use by QSIM, but in a more natural format. CONFIG and CC are both based on a component-connection ontology, so a mapping from CC's knowledge representation to CONFIG's knowledge representation was chosen as the focus of this study.

A mapping from CC to CONFIG was developed. Due to differences between the two programs, however, the mapping transforms some of the CC knowledge to CONFIG as documentation rather than as knowledge in a form useful to computation.

The study suggests that it may be worthwhile to pursue the mappings further. By implementing the mapping as a program, actual comparisons of computational efficiency and quality of results can be made between the QSIM and CONFIG programs. A secondary study may reveal that the results of the two programs augment one another, contradict one another, or differ only slightly. If the latter, the qualitative reasoning techniques may be compared in other areas, such as computational efficiency.

INTRODUCTION

Qualitative Reasoning

Quantitative reasoning is reasoning with precise, usually numeric, values. Applications such as expert systems, modeling, and simulation typically have been developed using quantitative reasoning. Due to the combinatorial and sometimes imprecise nature of the data in these applications, however, their success has been limited by the overwhelming amount of computation necessary to obtain the desired results. Qualitative reasoning can help ease this problem.

Qualitative reasoning can be used to reason over the same domains as quantitative reasoning, but a domain is represented as a (usually) small set of qualitative values rather than a large (and possibly infinite) set of quantitative values. This abstraction of a domain must be chosen carefully so that the set of qualitative values represents the important qualities of the domain while suppressing or ignoring other qualities. If the quantitative domain is continuous, the qualitative values chosen usually represent contiguous intervals of the continuous space. For example, a number line can be represented quantitatively by the real numbers; however, in some applications the same number line could be represented by the qualitative values "negative", "zero", and "positive". Since the number of elements in a qualitative domain is smaller than the number in the corresponding quantitative domain, combinatorial problems can be solved with less effort. A good introduction to qualitative reasoning in physical systems can be found in [6].

Although qualitative reasoning methods are still mostly experimental, their application to modeling, simulation, and analysis of physical systems holds great promise for use in fault diagnosis of those systems. Using qualitative reasoning techniques, fault diagnosis can be performed not only after a physical system has been built, but also during each phase of its design. The potential savings in time, effort, and money are enormous. Research groups, such as those headed by Forbus [2,4,7,8] and Kuipers [5,9,11,12], have developed experimental qualitative reasoning techniques and systems. A Johnson Space Center group headed by Dr. Jane Malin is developing a prototype qualitative reasoning program named CONFIG [13,14] for use in model-based fault diagnosis applications within the U.S. space program.

Goals and Rationale

These groups have taken (sometimes radically) different approaches to qualitative modeling, simulation, and analysis. Vigorous research is necessary to advance qualitative reasoning techniques to a level where they can be applied to solve a wide variety of real-world problems. However, sometimes the diverse methods utilized by different groups present problems: different knowledge representation schemes can make it difficult to share knowledge between groups, and incompatible reasoning techniques can make it

difficult to incorporate new techniques into existing programs. The thrust of this project is to study the former, and perhaps to gain insight into the latter.

The major goal of this project is to develop a mapping from the knowledge representations used by the Kuipers group (in the CC [9] and QSIM [5] systems) to that used by the Malin group (in the CONFIG system). A program could be developed from this mapping to mechanize the transformation of CC and/or QSIM knowledge to a form compatible with CONFIG. In addition to facilitating the sharing of physical systems models between research groups, such a program may serve as the basis for comparisons between reasoning methods used by the programs.

It is important that the mapping produced be correct and complete; i.e., all of the knowledge in the source representation should be transformed in such a way as to accurately retain its original meaning, and only that knowledge should be transformed (nothing should be added). This presents a problem when the target representation contains no counterpart for particular elements of knowledge embodied in the source representation. It was decided that these elements and all documentation elements would be transformed to documentation in the target representation, and all other knowledge would be transformed to its counterpart in the target representation.

CONFIG

General Description

CONFIG is a system that integrates object-oriented modeling, quantitative and qualitative mathematics, discrete event simulation, and digraph analysis to support the modeling, simulation, and analysis of physical systems. The initial application of CONFIG will be model-based fault diagnosis. CONFIG is based on a component-connection ontology [1,3]; i.e., a physical system is modeled as a collection of components and the connections between them. A component-connection model is extremely modular and, thus, can be modified easily. A CONFIG model can be viewed as a directed graph in which components, called devices, are the graph nodes and connections, called relations, are the graph edges. During simulation, changes in the value of device variables are propagated along the edges of the graph (i.e., through the component-to-component connections).

Knowledge Representation

Each device class is comprised of several elements (or parts), some of which may also be made up of parts. This arrangement of parts within parts is called a parts hierarchy. Figure 1 shows the parts hierarchy of a CONFIG device. The parts of a device are the internal variable clusters (VCs), port VCs, mode independent (MI) processes, a mode transition digraph (MTD), and a Device Control Digraph (DCD). In turn, the MTD is

made up of modes, and modes are composed of mode dependent (MD) processes and mode transition (MT) processes. These parts are detailed in [13] and [14].

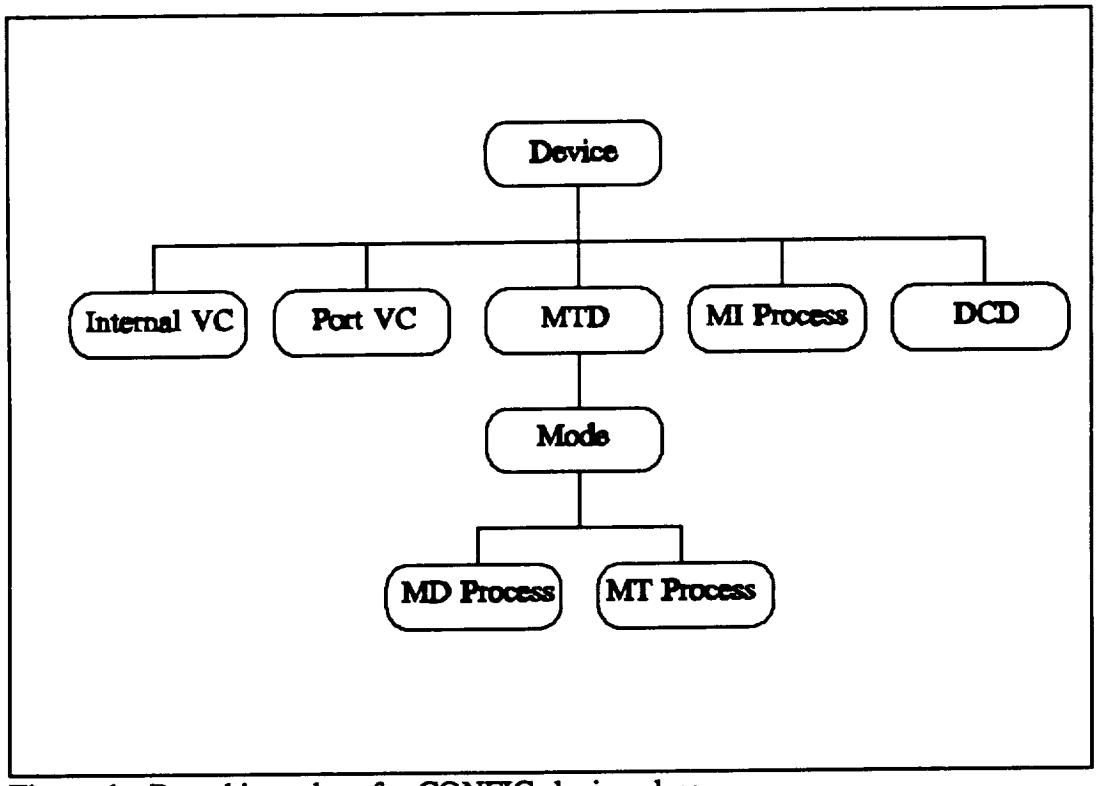


Figure 1.- Parts hierarchy of a CONFIG device class.

The purpose of a variable cluster (VC) is to aggregate logically related variables into a common data structure. The variables in the internal and port VCs contain state information about the device. An internal VC is similar to a private structure in a C++ class: it contains variables that are accessible only by the other parts of the device. A port VC, by contrast, is designed to be a device interface to other devices.

MI processes are one of three types of processes defined for use in discrete simulation (the others are described below). Semantically, a process has the general form

if *precondition* then *action* after *time-delay*.

If the precondition of a MI process becomes true during simulation, the action specified is placed on the simulation schedule to be executed after the time delay specified. The preconditions are checked each time the simulation clock is updated.

Conceptually, a device mode defines an operating region of the device and describes the behavior of the device when it is in that mode. By defining modes for each operating region of the device, the behavior of the device can be modeled completely. For

example, the modes for a switch device might be named "open", "closed", and "stuck" and describe how a switch behaves in each mode. The modes of an MTD describe a digraph in which the nodes are mode objects and the edges are MT processes. MT processes have the same general semantic form as MI processes. The MTD also contains a variable that indicates the current device mode. Each time the simulation clock is updated, the preconditions of the MT processes of the current mode are checked. If the precondition of one of the MT processes is true, then the corresponding action, which is a transition to another mode, takes place. The MTD also contains a collection of MD processes; these processes operate much like the MI processes, with the exception that their preconditions are checked only when the mode in which they are defined is the current mode.

The purpose of a DCD is to describe the submodel (i.e., subcomponent) structure within a device class.

A relation (i.e., an inter-device connection) can be thought of as a cable or pipe between a subset of the port VCs of two devices. The flow of data through the relation is generally one-way, though there are two-way relations (these can be viewed as two one-way relations). A relation may be defined to connect any two devices, and within these devices, any subset of their port VCs, and within the port VCs, any subset of the variables.

Simulation

CONFIG performs discrete event simulation. In discrete event simulation, events are assumed to occur at points in time rather than continuously over intervals of time. When the simulation begins, a sequence of events is placed in a schedule, ordered chronologically by their arrival times (i.e., the times at which the events will occur). The simulation clock is advanced to the arrival time of the first event, and the event takes place. As a consequence of event occurrence, other events may be scheduled. The simulation clock is advanced to the next event arrival time, and the process continues until the schedule is empty (or until a specified amount of clock time has passed). The purpose of using discrete event simulation is to simulate a numerical sample of the set of possible system state sequences, rather than generating all of them. (The method of generating all sequences is the basis of envisionment, a form of which is used in QSIM).

When CONFIG begins a simulation, each device is set to its specified initial mode. As the simulation progresses, the process conditions are checked, and if they are true, their actions are scheduled for execution. Data values are propagated from device port to device port through relations.

QSIM

General Description

QSIM [5] is a qualitative simulation system for model-based reasoning produced by the Kuipers research group at University of Texas at Austin. Although still considered an experimental system, QSIM has matured to the point that other research groups use it as a basis for their work. The modeling approach used in QSIM is radically different from that used in CONFIG. In QSIM, a system model is represented by a single qualitative differential equation (QDE)[1]. Contrast this to CONFIG, in which a system is modeled as a collection of components and connections. Furthermore, in QSIM it is possible to have multiple models of the same physical system, and to switch between models as needed during simulation. For example, the QSIM User's Guide [5] presents an example of modeling a bouncing ball; the system is modeled as an accelerating mass while in flight and as a spring when it rebounds the floor. QSIM's use of multiple models and model switching is similar in concept to CONFIG's use of device modes and mode transitions.

Knowledge Representation

A QDE specifies the structure of the system being modeled. A QDE is comprised of four major parts: a set of qualitative variables, a quantity space for each variable, a set of constraints among the variables, and a set of transition rules. Details about these may be found in [5].

A quantity space is a qualitative abstraction of a quantitative range. It consists of an ordered sequence of landmark values within the range; landmarks are borders between qualitative intervals. A QSIM user can specify any collection of landmarks as long as the collection includes the landmark 0 (zero).

The value of a qualitative variable is comprised of a qualitative magnitude, which is a value from the variable's quantity space, and the direction of the magnitude's change, which can either be decreasing, steady, or increasing.

A constraint specifies important relationships among the variables of the QDE. It consists of the constraint specification and an optional list of corresponding values. A constraint specification is a relation name and its variable arguments. A corresponding value is a list of variable values that further constrain the variables. For example,

((M+ A B) (0 0) (inf inf))

---

[1] A QSIM QDE construct is actually a set of qualitative differential equations, in the traditional sense of the term. However, to maintain consistency with QSIM documentation, the construct will continue to be referred to as a QDE for the remainder of this paper.

is a constraint in which "(M+ A B)" is a constraint specification, and "(0 0)" and "(inf inf)" are corresponding values. This constraint states that the relation M+ (positive monotonicity) holds between the variables "A" and "B". Further, the corresponding values state that when "A" is zero, "B" is zero, and when "A" is $+\infty$, "B" is $+\infty$.

Transition rules provide a mechanism for switching between models during simulation. Each transition rule consists of a condition and a QDE identifier; if the QDE containing the transition rule is active and the condition is true, the QDE specified in the rule becomes the active QDE.

Simulation

A simulation in QSIM differs drastically from one performed in CONFIG since QSIM uses a form of envisionment to perform simulation. Envisionment, as described in the section on CONFIG, is a process by which all possible system state sequences are generated. Beginning with the initial model in its initial state, QSIM generates a tree containing these state sequences, although QDE constraints and other constraining mechanisms are used to prune away those state sequences that violate constraints, thus saving time and computation. This tree is called a behavior tree. The user can request information about any path in the tree from QSIM. Even with pruning, though, building a behavior tree can be a computationally expensive process.

QSIM to CONFIG Mapping

The desirability of mapping QSIM knowledge to CONFIG was examined early in this study. As a result of this examination, it was decided that a QSIM-to-CONFIG mapping would not be developed. QSIM models all devices of the system in a single QDE. A QDE cannot reliably be separated into discrete devices. Thus, a QSIM model would be mapped to CONFIG as a single "mega-device". One of CONFIG's strong points is its component-connection representation. This representation is extremely modular; components and connections can easily be added or deleted. This modularity would be lost in the mapping.

CC

General Description

CC [9] is a preprocessor for QSIM; it takes CC input language as its input, and produces QSIM input language as its output. CC, like CONFIG, is based on a component-connection ontology. The CC input language is styled after that of VHDL [10], which separates a component's interface definition from its implementation definition so that multiple component implementations can be defined without the necessity of replicating the interface information. CC elements can be arranged into part

and inheritance hierarchies, even though QSIM does not support hierarchies; this is because CC "flattens out" the hierarchies.

Knowledge Representation

There are four major modeling elements in CC: quantity spaces, component interfaces, component implementations, and component configurations. Details about these elements can be found in [9].

Quantity spaces can appear either globally or within a component interface or implementation. The CC quantity space concept is an extension of the QSIM quantity space concept. Both contain an ordered list of qualitative landmark values, but a CC quantity space may also specify a parent quantity space and a list of conservation correspondences. A conservation correspondence is a list of qualitative values from the quantity space that sum to zero; thus, a conservation correspondence is a form of constraint on the quantity space values.

A CC component class definition is comprised of two parts: a component interface and a component implementation. A component interface is a high-level abstraction of a component that describes a family of component classes. The specific details of particular component classes are described by component implementations. Multiple implementations may be defined for a given interface. Each interface/implementation pair defines a component class. A component class with subcomponents is called a composed component class, and a component class without subcomponents is called a primitive component class.

The parts of a component class (i.e., an interface/implementation pair) include a knowledge domain name, terminals, component (local) variables, mode variables, subcomponents, and connections between the subcomponents.

The knowledge domain name is one of a list of predefined areas such as "hydraulic", "electrical", etc. A knowledge domain is a list of variable types commonly used in components within the domain; for example, the "hydraulic" domain contains variable types such as "flow" and "pressure". The authors of CC have taken a bond graph approach [15], so these domain-specific variable types simply are synonyms for a set of generic type names.

A terminal is an interface that can be connected to a terminal of another device. It contains a collection of terminal variables. Terminal and component variables are variables like those found in QSIM; i.e., their values are composed of a qualitative magnitude and a direction of change. A mode variable is a QSIM variable and a collection of condition/value pairs. During simulation, the conditions are checked, and if any is found to be true, its associated value is assigned as the value of the mode variable.

The final major CC element, the component configuration, might be called a "composed component macro". Its purpose is to create similar composed component class definitions from a single component class that has some subcomponent information omitted. The component configuration does this by supplying the missing subcomponent information. For example, assume that there exists a composed component class named "Black Box" with three subcomponents for which implementations (and perhaps interfaces) have not been specified. Any number of similar "Black Box-like" component classes can be created by defining component configurations that specify different subcomponent class information for the three subcomponents.

CC does not appear to support QSIM model switching. Although multiple QDE's can be generated, the CC User's Guide does not document any way of generating the QDE transition relations needed to switch between the models during simulation.

CC-TO-CONFIG

Order of Mapping

The order in which CC elements are mapped to CONFIG elements is bottom-up. That is, a CC construct is converted only if all of its subparts have already been converted, or if it contains no subparts. In this way, a given construct can be transformed in a single pass, whereas with a top-down approach multiple passes might be required to complete the transformation of a given construct. This leads to the following transformation order:

1) Global quantity spaces with no parent specified;
2) Global quantity spaces whose parent has already been transformed;
3) Primitive component classes;
4) Component configurations are transformed to CC composed component classes; and
5) Composed component classes whose subcomponents have already been transformed.

Note that step 4 is a CC-to-CC transformation, not a CC-to-CONFIG transformation. The purpose of this step is to convert component configurations to a form that is easier to transform to CONFIG.

Mapping Quantity Spaces

Quantity spaces, regardless of whether they appear globally or within a component class definition, most closely resemble qualitative data types in the CONFIG process specification language. The main elements of a quantity space are an ordered set of landmark values and a set of conservation correspondences on those values. A CONFIG qualitative data type consists of an unordered set of qualitative interval names and a collection of operations on those qualitative names. Transformation of the landmark values to interval names is straightforward; since landmarks represent the boundaries of qualitative intervals, interval names will be generated by joining adjacent landmark names. For example, the set of landmark values

(L1 L2 L3 ...)

can be transformed to the set of interval names

(L1_L2 L2_L3 ....).

The transformation of conservation correspondences to operations on the newly generated qualitative type is complicated by the fact that the conservation correspondences are operations on landmarks rather than on intervals. However, this can be overcome by generating additional interval names that correspond to the intervals between 0 (which is a landmark in every quantity space) and the landmarks in the conservation correspondences. Then the sum of these intervals can be viewed as zero, and these interval operations can be transformed to CONFIG qualitative data type operations. For example, the quantity space,

(minf a b 0 c d inf) (0 0) (minf inf) (b c d))

would be transformed to the qualitative type

(minf_a a_b b_0 0_c c_d d_inf 0_0 minf_0 0_inf b_0 0_c 0_d)

and the transformed conservation correspondences (i.e., those intervals whose sum is the interval 0_0) would be (in functional form):

(SUMS-TO-ZERO 0_0 0_0)
(SUMS-TO-ZERO minf_0 0_inf)
(SUMS-TO-ZERO b_0 0_c 0_d)).

In addition, the original list of landmark values will be copied to CONFIG to document the original landmark ordering.

Quantity Spaces with Inheritance

Communication with Dr. Kuipers about the inheritance mechanism for quantity spaces indicates that there are still many unanswered questions about how the mechanism will operate. Thus, it was decided to take the view that the landmarks and conservation correspondences of parent quantity spaces are recursively inherited by child quantity spaces. Thus, the transformation of a quantity space whose parent quantity space has already been transformed is achieved by taking the union of the parent's interval set and the child's interval set, and the union of the parent's operation set and the child's operation set. This view assumes, however, that particular landmarks have the same qualitative meaning at each level of inheritance.

Mapping Primitive Components

Primitive components are component classes with no subcomponents. They will be transformed to a CONFIG device class. This transformation involves many elements, which are described below.

The implementation and interface of the component class contain descriptive text strings. These will be included as documentation in the CONFIG device class. Other items transformed as documentation will be the component class knowledge domain name and component parameters.

CC terminals and CONFIG port variable clusters serve the same purpose, and each is comprised of a collection of variables. Thus each terminal will be transformed to a CONFIG port variable cluster class, and the CONFIG device class will contain an instance of each variable cluster class as a port.

In a similar fashion, the component (local) variables will collectively be transformed to a CONFIG variable cluster class, and the CONFIG device class will contain an instance of this variable cluster class as an internal variable cluster. CC terminal and component variables may also have several options specified, such as "display" and "no-new-landmarks"; these are operational flags that have no counterpart in CONFIG, and so will be copied as documentation for the variable.

A CC mode variable is a variable and a collection of condition/value pairs. If one of the conditions is true, the variable takes on the corresponding value. Mode variables have no relationship to modes in a CONFIG device. Their variables will collectively be transformed to a CONFIG variable cluster class, and the CONFIG device class will contain an instance of the variable cluster class as an internal variable cluster. Their condition/value pairs will be transformed to MI processes.

CONFIG does not use constraint propagation. Thus, transformation of CC constraints to CONFIG may have limited utility. The relations specified in most CC constraints have predefined meanings; for example, M+ is a relation that between two variables, one of which affects the value of the other in a monotonically positive manner. The predefined meanings will not be translated, but the list of corresponding values in the constraint could be transformed to CONFIG process language operations on the types of the constraint arguments.

CC constraints define the value of one variable as a function of another; i.e., whenever the domain variable value changes, the range variable value changes according to a function of the domain variable. This relationship could be transformed to CONFIG MI or MD processes in which the change of domain variable from one qualitative value to another triggers an action that changes the range variable value accordingly. However, since only the name of the function rather than the function itself is encoded in the

constraint, the transformation from constraint to MI or MD processes would have to be performed by the human modeler. If the modeler chooses not to perform this task, then the constraints would be transformed to CONFIG device class documentation.

Mapping Composed Components

The conversions described for primitive component classes also hold for composed component classes. There are two additional elements that must be converted for composed components: subcomponents and connections between subcomponents. By taking a bottom-up approach, the subcomponents already will have been converted, so instances of the subcomponent device classes can be included in the composed device class DCD. Subcomponent connections will be transformed to CONFIG relations, since both describe device-to-device connections, and will also be included in the DCD.

CONCLUSIONS AND FUTURE WORK

A CC-to-CONFIG knowledge representation mapping has been described. Although it is complete (i.e., it describes the transformation of all CC knowledge to the CONFIG representation), several elements are translated as documentation, due to differences in CC/QSIM and CONFIG. These gaps represent areas of further study for potential extension of CONFIG's capabilities. The mapping also highlights the need for further studies concerning mappings from other qualitative reasoning programs to CONFIG, and also the mapping of CONFIG representations to other programs. For example, even though a mapping from QSIM to CONFIG is not desirable, a mapping from CONFIG to QSIM might be. In summary, the current study has revealed more questions to be investigated.

A good starting point for further study would be the implementation of the mapping described here. Doing so will open the doors for empirical comparisons of CONFIG to QSIM. It is important to know if the results achieved by CONFIG are complimentary, contradictory, or essentially the same as the results of other programs. If complimentary, the programs could be used together to get better results. If contradictory, why? If essentially the same, comparisons could be made to find out if one program has a better computational efficiency than the others.

REFERENCES

1. Abelson, H.; Sussman, G.J.: The Structure and Interpretation of Computer Programs, MIT Press, 1985.

2. DeCoste, D.: Dynamic Across-Time Measurement Interpretation. *Artificial Intelligence*, vol. 51, 1991, pp. 273-341.

3. de Kleer, J.; and Brown, J.S.: A Qualitative Physics Based on Confluences. Qualitative Reasoning About Physical Systems, Daniel G. Bobrow, ed., MIT Press, 1985, pp. 7-83.

4. Falkenhainer, B.; and Forbus, K.D.: Compositional Modeling: Finding the Right Model for the Job. *Artificial Intelligence*, vol. 51, pp. 95-143.

5. Farquhar, A.; Kuipers, B.; Rickel, J.; Throop, D.; and The Qualitative Reasoning Group: QSIM: The Program and its Use. Draft. Department of Computer Sciences, University of Texas at Austin, Austin, Texas, 1992.

6. Forbus, K.D.: Qualitative Physics: Past, Present, and Future. Exploring Artificial Intelligence: Survey Talks from the National Conferences on Artificial Intelligence, H.E. Shrobe, ed., Morgan Kaufmann, 1988, pp. 239-296.

7. Forbus, K.; and Falkenhainer, B.: Self-Explanatory Simulations: Scaling Up to Large Models. Proceedings of the Tenth National Conference on Artificial Intelligence, AAAI Press, 1992.

8. Forbus, K.D.; Nielsen, P; and Faltings, B.: Qualitative Spatial Reasoning: The CLOCK Project. *Artificial Intelligence*, vol. 51, 1991, pp. 417-471.

9. Franke, D.W.; and Dvorak, D.L.: CC: Component Connection Models for Qualitative Simulation, A User's Guide. Draft. Department of Computer Sciences, University of Texas at Austin, Austin, Texas, 1992.

10. IEEE Standard VHDL Language Reference Manual. IEEE Std. 1076-1987.

11. Kuipers, B.: Component-Connection Models. Draft. University of Texas at Austin, 1992.

12. Kuipers, B.J.; Chiu, C.; Dalle Molle, D.T.; and Throop, D.R.: Higher-Order Derivative Constraints in Qualitative Simulation. *Artificial Intelligence*, vol. 51, 1991, pp. 343-379.

13. Malin, J.T.; and The CONFIG Design Team: CONFIG User's Guide. Draft. Intelligent Systems Branch, Automation and Robotics Division, Johnson Space Center, 1990.

14. Malin, J.T.; Basham, B.D.; and Harris, R.A.: Use of Qualitative Models in Discrete Event Simulation to Analyze Malfunctions in Processing Systems. Artificial Intelligence in Process Engineering, Academic Press, 1990, pp. 37-79.

15. Rosenberg, R.C.; Karnopp, D.C.: Introduction to Physical System Dynamics. McGraw-Hill, 1983.

N93-26073

INVESTIGATION INTO THE COMMON MODE REJECTION RATIO OF THE PHYSIOLOGICAL SIGNAL CONDITIONER CIRCUIT

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1992

Johnson Space Center

Prepared By:	Edward M. O'Brien, Ph.D.
Academic Rank:	Associate Professor
University & Department:	Mercer University Electrical and Computer Engineering Department Macon, Georgia 31207
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Life Sciences Project
Branch:	Project Engineering
JSC Colleague:	Donald J. Stilwell
Date Submitted:	October 30, 1992
Contract Number:	NGT-44-005-803

ABSTRACT

The common mode rejection ratio (CMRR) of the single operational amplifier (op amp) differential amplifier and of the three operational amplifier differential amplifier was investigated. The three op amp differential amplifier circuit is used in the signal conditioner circuit which amplifies signals such as the electromyograph or electrocardiogram. The investigation confirmed via SPICE modeling what has been observed by others in the recent literature that the CMRR for the circuit can be maximized without precision resistor values or precisely matched op amps. This can be done if one resistor in the final stage can be adjusted either by a potentiometer or by laser trimming in the case of hybrid circuit fabrication.

INTRODUCTION

The physiological signal conditioning circuit is used to amplify various signals such as the electrocardiogram (ECG), the electromyogram (EMG), the electrooculogram (EOG), and the electroencephalogram (EEG). The electrodes used to record these signals will often see a voltage that is common to them. This will happen in the case of EEG electrodes that will also see a voltage due to the ECG. Also, 60 Hz noise is almost always capacitively coupled to the individual being monitored. Hence a differential amplifier is used to subtract out the common voltage and amplify the difference voltage between the electrodes. A measure of how well an amplifier amplifies the difference signal and attenuates the common signal is called the common mode rejection ratio (CMRR). It is important to keep this value as high as possible since some of the signals are very low in amplitude. Thus they are susceptible to being obscured by any voltage that may be in common with the transducing electrodes. This paper investigates the common mode rejection ratio of the one and three operational amplifier (op amp) differential amplifiers.

GENERAL BACKGROUND ON DIFFERENTIAL AMPLIFIERS

A detailed presentation into the theory of cascaded differential stages and differential amplifiers has been presented by Pallás-Areny and Webster (1991 a and b) and will be briefly discussed here. Figure 1 shows a single differential stage. It is assumed that each stage has two inputs and two outputs. These stages can be cascaded. The relationship between the inputs and outputs are governed by the following four terms: G_{DD}, Equation (1), is the differential mode gain assuming that there is no common mode voltage on the inputs. GCC, Equation (2), is the common mode gain assuming that Vo_1 and Vi_2 are equal. G_{DC}, Equation (3), is the common mode input to difference mode output gain. G_{CD}, Equation (4), is the difference mode input to common mode output gain. Two additional terms, C_i and D_i, are defined by Equations (5) and (6) respectively. The actual CMRR of each stage is defined in terms of the C and D values for the current stage and the preceding stages. The total $CMRR_T$ for n cascaded stages is given by Equation (7) and is the reciprocal of the individual stage CMRR's. This seems counter to the way most individuals think of cascaded stage CMRR's as being multiplicative. This intuitive concept can however be seen in Equation (8) which shows that the CMRR of each stage is the product of its own C_i term and the previous stages' D terms. A short discussion of the application of this formula to the single and three op amp differential amplifiers will be presented.

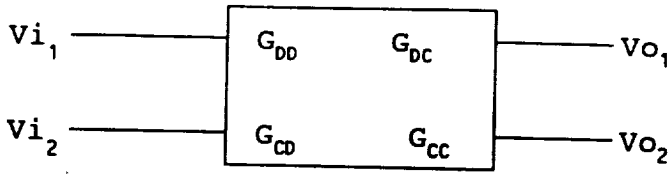


Figure 1. Two input and two output diagram of a differential amplifier.

$$G_{DD}=\frac{VO_1-VO_2}{Vi_1-Vi_2} \tag{1}$$

$$V_{CC}=\frac{(VO_1+VO_2)/2}{(Vi_1+Vi_2)/2} \tag{2}$$

$$G_{DC}=\frac{VO_1-VO_2}{(Vi_1+Vi_2)/2} \tag{3}$$

$$G_{CD}=\frac{(VO_1+VO_2)/2}{Vi_1-Vi_2} \tag{4}$$

$$C_i=G_{DD_i}/G_{DC_i} \tag{5}$$

$$D_i=G_{DD_i}/G_{CC_i} \tag{6}$$

$$\frac{1}{CMRR_T}=\sum_{i=1}^{n}\frac{1}{CMRR_i} \tag{7}$$

$$CMRR_i=C_i\prod_{1}^{i-1}D_i \tag{8}$$

SINGLE STAGE DIFFERENTIAL AMPLIFIER

In this section the CMRR of the one op amp differential amplifier is examined. Figure 2 shows the complete schematic of all the possible circuit configurations investigated in this paper. Operational amplifier, OA1, and resistors, R1-R4, make up the single op amp differential amplifier. This is really a two stage cascaded differential amplifier. The op amp itself is the last stage and the resistive network is the first stage. The inputs for the resistive stage are V_1 and V_2 and the outputs are at the inverting and noninverting terminal of OA1. The C term for the resistive stage can be calculated to be as shown in Equation (9). The D term for this stage is 1. The C term for the OA1 is just the CMRR of the amplifier as presented in the specifications. For the LT1078AC

$$C=\frac{(2R_2R_4+R_2R_3+R_3R_4)/2}{R_2R_3-R_1R_4} \qquad (9)$$

micropower op amp the typical value for the CMRR is 110 Db with a minimum value of97 Db at 25° centigrade. This is the low frequency value. The C for this op amp actually starts to roll off with frequency with a 3 Db corner frequency of about 100 Hz. Applying Equation (8) the actual CMRR for the op amp is the product of its C term and the prior stage's D term. This yields a value 106 Db *1 or just 106 Db. Applying Equation (7) the reciprocal of the $CMRR_T$ for the one op amp differential amplifier is equal to the sum of inverses of the CMRR's for the resistive stage, i.e., the inverse of Equation (9), plus the inverse of the op amp CMRR. This would lead one to think that the $CMRR_T$ would always be less than that of lowest CMRR of the two stages just like the effective resistance of parallel resistors. However, by proper choice of resistor values, Equation (9) can actually be negative and thus, in theory, can exactly cancel the CMRR of the OA1 stage. This would result in an infinite $CMRR_T$ at least at frequencies below 100 Hz.

A PSPICE model of the one op amp differential amplifier was examined for the sensitivity of the output to the resistors R1-R4. The normalized sensitivities in units of volts/percent are shown in Table I. The common mode input voltage was 8 V dc. The nominal values of the resistors are also shown in the table. With the values shown, the stage has a differential gain of 10. A ten percent change in R4 would result in an output voltage change of .727 V. It should be noted that the magnitudes of the resistor sensitivity are the same and therefore it would not matter which one was varied. If the resistor values are exactly as indicated in Table I, the CMRR of the resistor stage would

be infinite and therefore the $CMRR_T$ for both stages would be that of the op amp used. If an LT0178 were used that had a CMRR of 106 Db, and if R1-R3 were the exactly as in Table I, then R4 would need to have a value of 199988.98 for the resistor stage to have a CMRR of -106.004 Db. This would cancel the CMRR of the op amp giving a $CMRR_T$ approximately infinite. If adjustment of the amplifier is possible after fabrication then precise resistors need not be used. Low tolerance resistors together with a fixed resistor and a trimmer pot can be used to effectively increase the $CMRR_T$ of the stage. Laser trimming of resistors can also be done in hybrid fabrication.

Table I

Resistor	Value	Normalized Sensitivity (V/percent)
R1	20 kΩ	7.27E-02
R2	200 kΩ	-7.27E-02
R3	20 kΩ	-7.27E-02
R4	200 kΩ	7.27E-02

In the laboratory, the one op amp differential amplifier was constructed using a LT1078 op amp. The values of the resistors are shown in Table II. A 5 V dc and a 5 V ac (1 Hz - 10.0 Khz) source was applied to the two inputs. A trimmer potentiometer comprising part of R4 was adjusted to achieve a minimum output voltage. The resulting value of R4 is shown in the table. The output voltage was minimized until obscured by .5 mV of noise on the output. Thus the CMRR of the stage was better than 100 Db. Equation (9) can be used to calculate the CMRR of the resistor stage and yields a value of -97.556 Db. This is certainly within the range of CMRR's specified for this op amp. The output voltage remained constant up to 1000 Hz and then started dropping. This indicates that the 3 Db point for this op amp was better than the 100 Hz which was indicated in the specification. At 10.0 Khz the CMRR of the stage dropped to 83.3 Db.

Pallás-Areny and Webster (1990) used a LM741C op amp with a specified 90 Db CMRR and obtained a $CMRR_T$ at 10 Hz and below of 126 Db (which was the limit of their measurement capabilities). This was done with nonprecision resistors and a trim pot. As expected, if the op amp was exchanged for another the potentiometer had to be readjusted to minimize $CMRR_T$.

Table II Laboratory resistor values for one op amp differential amplifier.

Resistor	Value
R1	20.0184 kΩ
R2	191.194 kΩ
R3	20.0079 kΩ
R4	191.067 kΩ

Table III Resistor values and sensitivities for the 3 op amp differential amplifier.

Resistor	Value	Normalized Sensitivity (V/percent)
R1	20 kΩ	7.27E-02
R2	200 kΩ	-7.27E-02
R3	20 kΩ	-7.27E-02
R4	200 kΩ	7.27E-02
R5	2 kΩ	-8.70E-10
R6	100 kΩ	5.51E-05
R7	100 kΩ	-5.51E-05

THREE OP AMP DIFFERENTIAL AMPLIFIER

The three op amp differential amplifier is shown in Figure 2 and comprises op amps OA1-OA3 and resistors R1-R7. For the present discussion OA2, OA3 and R5-R7 will be considered as one stage. The common mode rejection ration for this first stage, $CMRR_F$, [i.e., its C term as defined in Equation (5)] has been developed by Pallás-Areny and Webster and just the results will be reproduced here in Equation (10). In this equation the A_d's and the A_c's refer to the difference gain and common gains of the OA2 and OA3. It should be noted that R5-R7 are not involved with the CMRR of this stage. The following will indicate why this is so.

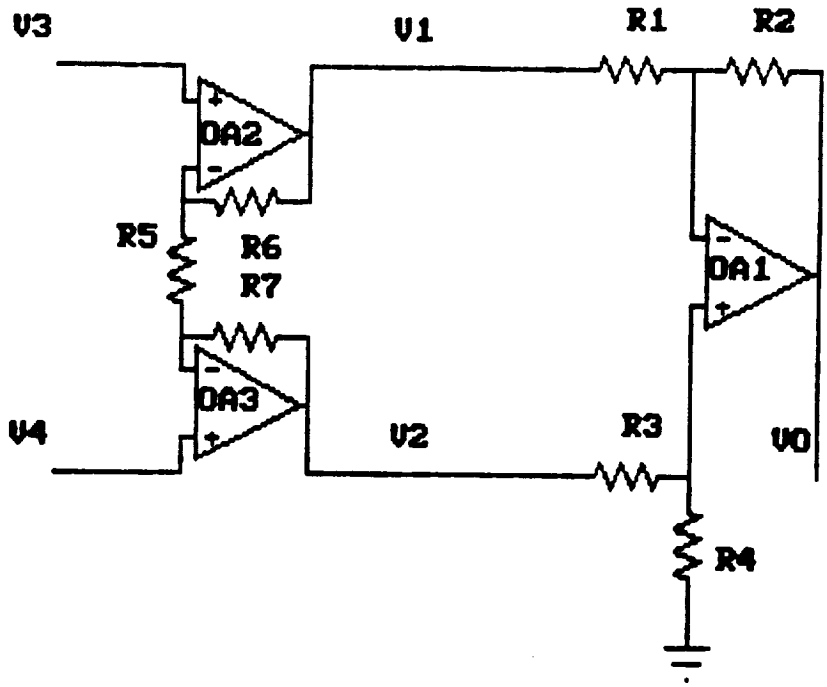


Figure 2. One and three op amp differential amplifier.

The sensitivity of the output to all resistors, R1-R7 was investigated with a PSPICE program. Again, the input common mode voltage was 8 V and the sensitivities are reported in terms of volts/percent. The results of the analysis are shown in Table III. The

$$\frac{1}{2}\,\frac{A_{d2}+A_{c2}/2+A_{d3}+A_{c3}/2+2A_{d2}A_{d3}-A_{c2}A_{c3}/2}{A_{d3}+A_{c3}/2-A_{d2}-A_{c2}/2+A_{d2}A_{c3}-A_{d3}A_{c2}} \qquad (10)$$

sensitivities of R1-R4 are the same as those found for the 1 op amp differential amplifier. The sensitivities of R6 and R7 are the same except for a negative sign and these sensitivities are approximately 1300 times less than for R1-R4. Thus they affect the common mode output very little. R5 has virtually no effect on the common mode voltage as seen by its very low sensitivity. From these results the maximum change in the common mode voltage gain is associated with resistors R1-R4.

From Equation (10) it is evident that for high a high $CMRR_F$ the two input op amps OA2 and OA3 must have both their common mode gains and difference mode gains matched. It is not enough to just have equal CMRR's for each.

The D term for this first stage, as defined in Equation (6), is given by Equation (11) and it can be found in many books on electronics (e.g., Horowitz and Hill, 1989 p. 425). This assumes that R6 and R7 are equal. The term $CMRR_s$ will be used for the common mode rejection ratio of the second stage of the three op amp differential amplifier. It will now consist of the CMRR found for this stage times the D term for the preceding stage as indicated by Equation (8).

$$D=1+\frac{2R6}{R5} \qquad (11)$$

Therefore, the $CMRR_T$ for the three op amp differential amplifier is given by equation (7) and shown more explicitly in Equation (12). If $CMRR_F$ and $CMRR_S$ have equal sign then $CMRR_T$ will be less that the lowest of the individual stage's CMRR. However, again by

$$\frac{1}{CMRR_T}=\frac{1}{CMRR_F}+\frac{1}{CMRR_S} \qquad (12)$$

adjusting one of the resistors R1-R4, $CMRR_S$ can be either positive or negative. Thus the total CMRR of the whole differential amplifier circuit can be in theory, made close to infinity. PSPICE modeling of this circuit verified this observation. Adjustment of R4 would allow the $CMRR_T$ to be driven arbitrarily close to infinity even when the A_d's and A_c's of OA2 and OA3 were greatly mismatched. Adjustment of R4 can even compensate for a mismatch of R6 and R7.

SUMMARY

In summary, if no adjustment can be made of the resistor values in the one or three op amp differential amplifier, then all resistor values should be precise and the two op amps in the first stage of the three op amp differential amplifier must have their common mode gains and their difference mode gains matched. Also each op amp should have a high CMRR. However, if one of the resistors in the last stage can be adjusted to null the overall CMRR of the amplifier then precision resistors and exactly matched operational amplifiers need not be used.

REFERENCES

Horowitz, P. and W. Hill. 1989. The Art of Electronics. Cambridge University Press, Cambridge. 1125 pp.

Pallás-Areny, R. and J.G. Webster. 1990. Composite Instrumentation Amplifier for Biopotentials. An. of Bio. Eng., vol 18, 251-262.

Pallás-Areny, R. and J.G. Webster. 1991a. Common Mode Rejection Ratio in Differential Amplifiers. IEEE Trans. Inst. and Meas., vol 40, No. 4, 669-676.

Pallás-Areny, R. and J.G. Webster. 1991b. Common Mode Rejection Ratio for Cascaded Differential Amplifier Stages. IEEE Trans. Inst. and Meas., vol 40, No. 4, 677-681.

N93-26074

IMPLEMENTATION OF QUALITY IMPROVEMENT TECHNIQUES FOR MANAGEMENT & TECHNICAL PROCESSES IN THE ACRV PROJECT

Final Report
NASA/ASEE Summer Faculty Fellowship Program--1992

Johnson Space Center

Prepared by:	Laura B. Raiman, Ph.D.
Academic Rank:	Assistant Professor
University & Department:	Pennsylvania State University Department of Industrial & Systems Engineering University Park, PA 16802
NASA/JSC	
Directorate:	New Initiatives Office
Division:	ACRV Project Office
Branch:	n/a
JSC Colleague:	Brian K. Kelly
Date Submitted:	August 7, 1992
Contract Number:	NGT-44-005-803

ABSTRACT

Total Quality Management (TQM) is a cooperative form of doing business that relies on the talents of everyone in an organization to continually improve quality and productivity, using teams and an assortment of statistical and measurement tools. The objective of the activities described in this paper was to implement effective improvement tools and techniques in order to build work processes which support good management and technical decisions and actions which are crucial to the success of the ACRV project. The objectives were met by applications in both the technical and management areas. The management applications involved initiating focused continuous improvement projects with widespread team membership. The technical applications involved applying proven statistical tools and techniques to the technical issues associated with the ACRV Project.

Specific activities related to the objective included working with a support contractor team to improve support processes, examining processes involved in international activities, a series of tutorials presented to the New Initiatives Office and support contractors, a briefing to NIO managers, and work with the NIO Q+ Team. On the technical side, work included analyzing data from the large-scale W.A.T.E.R. test, landing mode trade analyses, and targeting probability calculations. The results of these efforts will help to develop a disciplined, ongoing process for producing fundamental decisions and actions that shape and guide the ACRV organization.

INTRODUCTION

Total Quality Management (TQM) is a cooperative form of doing business that relies on the talents of everyone in an organization to continually improve quality and productivity, using teams and an assortment of statistical and measurement tools. JSC management has made a commitment to understanding TQM techniques, and to implementation within JSC organizations. The Assured Crew Return Vehicle (ACRV) Project Office began to investigate and utilize the applications and benefits of continuous quality improvement techniques during the late spring and summer of 1991. As the ACRV project continues to move through its various stages of development, it remains vital that effectiveness and efficiency be maintained in order to provide the Space Station Freedom (SSF) crew an affordable, on-time assured return to Earth. The challenge is magnified by an increasing pressure to adhere, if not reduce, schedules and cost while meeting strict technical objectives. The introduction of ESA and NPO Energia into the picture has created a whole new set of issues and processes that need to be addressed. An additional factor for the success of the ACRV is attaining the maximum benefit from the resources applied to the project.

This paper documents the work performed over a 10 week period which involved examining, measuring and assessing processes being utilized by the ACRV Project teams. A major portion of this work involved determining specifically where certain statistical tools would be appropriate, and demonstrating their application. Major areas of interest included ACRV testing techniques, probability of mission success (POMS), and support contractor processes. The results from this work are aiding the ACRV office in implementing a disciplined, ongoing process for producing fundamental decisions and actions that shape and guide the organization. The resulting activities of the ACRV Project teams will serve as models for other organizations which are attempting to use TQM effectively to anticipate and respond to changing environments.

METHODS IMPLEMENTED FOR IMPROVEMENT ACTIVITIES

The activities conducted over the summer were supported by a series of four tutorials presented by the Summer Faculty Fellow (SFF) to the New Initiatives Office (NIO) and associated support contractors, a briefing to NIO management, and interaction with the NIO Q+ Team, and the ACRV and NIO manager. Direct application of numerous quality improvement techniques was accomplished in conjunction with the administrative support contractor personnel in the ACRV office. The SSF also encouraged and supported direct use of concepts and tools presented in the tutorials. Throughout the tutorials, participants were asked to brainstorm and discuss related topics, particularly with regard to implications and applications within their own organizations. All tutorials were videotaped, and copies of the tapes and the viewgraphs used may be obtained through the NIO Library in the NOVA building. The four tutorials presented are described below.

1. *Implementing A Total Quality System: Process Improvement and The Role of Deming's Points*

This tutorial presented the basic concepts included in a total quality system, and discussed the elements needed for implementing a quality improvement system. Topics included leadership involvement, team structure, team facilitation, quality training, and quality planning. The first 7 of Deming's 14 points were presented, including discussions relating the implications of the points to the attendees' organizations. In the Deming context, a model of quality was presented as it relates to a business strategy. Time was also spent discussing a methodology for process improvement, and how to make implementation of the various steps happen.

2. *The Implications of Variation on Continuous Improvement and Deming Points 8-14*

It is vital that all team members understand some of the basic statistical concepts needed to interpret variation. This tutorial presented an overview of the impact of variation, how it can be measured, and how it can be monitored for valuable results. Deming's Red Bead experiment was performed, and the results were

analyzed through the use of control charts and discussion. The necessity to make a distinction between patterns of variation was stressed in order to minimize the losses resulting from the misinterpretation of the patterns. These losses can be minimized by understanding that variation can be caused by either common or special causes, by knowing how to determine whether a system is stable or not, and by basing action on this analysis. Some methods for creating control charts were presented, with significant time spent on discussing the interpretation of control charts.

3. *A Toolbox for Quality Improvement*

This tutorial began with a re-introduction to the role of problem solving, and a strategy for process improvement. The roles and responsibilities of team members was presented and discussed. The remainder of the tutorial presented a wide range of tools which may be used by teams in continuous improvement efforts. Effective use of brainstorming and its relations to team activities was presented. The various tools presented included:

Seven Old Tools: flowcharts, cause and effect diagrams, pareto diagrams, checksheets, histograms, scatter diagrams, and control charts.

Seven New Tools: Affinity Diagrams, Interrelationship Diagrams, Tree Diagrams, Matrix Diagrams, Matrix Analysis, PDPC, and Arrow Diagrams.

Quality Function Deployment

Brainstorming

4. *"ReInventing Government"*

This tutorial presented a discussion of the recently published book of the same title. Discussions included a relation of Deming's 14 points in light of relationships to government organizations. Significant time was spent discussing examples of successful applications of continuous improvement efforts in a widespread assortment of governmental organizations.

IMPLEMENTATION OF QUALITY IMPROVEMENT TECHNIQUES

One highly focused team improvement activity was accomplished over the 10 week period. This involved the work being done by the ACRV support contractor - Eagle Engineering. An office of three people performs all of the administrative support for the ACRV project office. These activities include typing letters, typing presentations, copying, preparing viewgraphs, sending and receiving faxes, scheduling telecons, keeping files and retrieving documents, maintaining an office calendar, and numerous other activities. In preparation for an expansion of the activities of the project office, this team wanted to understand and improve their processes before the introduction of international activities completely confounded the support activities.

Throughout the summer, data was collected by tracking many of the activities performed in the support office. A log was kept of copying performed, copier breakdowns, and fax breakdowns. An "Administrative Support WORK REQUEST" form was also developed and tracked throughout the entire 10 week period. This form helped keep track of how long it took jobs to go through the office, who was requesting work, what type of work was being requested, and barriers that were encountered along the way. Two surveys were also given to customers during two weeks of Configuration Review meetings that were supported by this same team. Towards the end of the tracking period, all results were collected and analyzed. Through the use of control charts, bar charts, pareto diagrams, and flowcharts, the support personnel were able to understand the current state of their processes, and determine methods for improvement.

The results of these activities helped the support contractor team to improve the processes that they dealt with daily in supporting the work of the project office. The results of this activity also helped other project office personnel understand the significant amount of activity necessary in project support, and how their help was needed to maintain effective support on the part of the contractor. This eliminated some of the barriers between management and personnel, and between the contractor and the project office personnel.

ENGINEERING ANALYSIS ISSUES

Three major activities took place in the area of statistical analysis. These included the analysis of W.A.T.E.R. test results, aiding the Landing Mode Decision Team in decision analysis, and further calculation of targeting probabilities.

W.A.T.E.R. Test Results

This activity involved analyzing results from a test which the SFF helped design during the previous summer. The Wave Analysis & Test of Extraction Requirements (WATER) Test was a large-scale testing effort which focused on the post-landing phase of the ACRV mission, specifically a water-landing vehicle. The development of a full scale, generic ACRV mockup was the centerpiece of the crew egress evaluation. A mockup to simulate the water dynamics of both the Apollo and a NASA study vehicle concept, SCRAM. An entire week of unmanned testing took place in May, prior to the SFF's arrival. The SFF was then able to attend a portion of the manned testing which occurred throughout the first week in June.

All of the results of these testing efforts had to be analyzed using experimental design analysis techniques. This included synthesizing numerical, quantitative, results with verbal, qualitative input offered by the Search and Rescue (SAR) personnel. Response variables such as flotation attitude, pitch amplitude, yaw rate, and wave run-up had to be analyzed relative to input variables such as vehicle configuration, wave state, hatch location and crew composition.

Landing Mode Decision Team

This effort involved the use of a tool called the Analytic Hierarchy Process (AHP) to help the team come to a recommendation regarding the required capability of an ACRV to land on land or water. The team had already gone through a process of defining criteria/objectives to make the decision, definitions of these criteria, and measures of effectiveness for each criteria. Throughout the summer, for each measure of effectiveness, a score was determined for one of five landing modes: land only, water only,

land and water, land primary with water back-up, and water primary with land back-up. The SFF then helped the team in the use of AHP to rank by importance the measures of effectiveness, rank by importance the mandatory and highly desirable criteria, and then to combine all of these rankings together into a final recommendation. The results of this work will help provide the direction for future development efforts of the ACRV.

Targeting Probabilities

This effort involved calculating targeting probabilities for landing the ACRV on land sites throughout the world, and the related probability of mission success. Initial engineering studies were performed to estimate the targeting accuracy to a landing point of a range of ACRV configurations dues to navigation, entry guidance, and parachute drift dispersions. Information was used to generate a two-dimensional probability distribution which characterized the probability of landing over a given region. Using information obtained on obstacle coverage at two specified landing sites, a probability distribution for the obstacles was also generated. When these two probability footprints were combined, the results provide the capability of estimating the probability of the ACRV hitting an obstacle upon landing. This measure is one of a number of criteria that play into the overall probability of mission success. This also will play a role in determining the landing mode decision described previously.

JSC CENTER ACTIVITIES

Along with the activities described above, throughout the summer the SFF also spent some time working with groups outside of the ACRV Project Office. These activities included meeting with the JSC TQM Steering Committee, presenting a TQM tutorial to the Budgeting office, and talking to the Manager of Training and Development regarding center-wide training in quality improvement methods and tools. All of these activities grew out of the initial work and tutorials provided to the ACRV project team and NIO personnel, and will help support the Center's efforts in implementing and sustaining continuous improvement.

CONCLUSIONS

During the course of the summer's activities, numerous people in NIO and the ACRV project office have gained a better understanding of TQM - not just as a philosophy, but also as knowledge of the tools available to support the philosophy. This has resulted in more people thinking about how to use the available tools, and possible applications in the work and process of NIO activities. Members of the ACRV office, and NIO, expressed a strong interest in finding more applications, and continuing the quality improvement activities so that knowledge of the tools is not lost. For the administrative support team, significant improvements have been made in the processes evaluated by the team.

More people are now looking at how to implement TQM in their own organizations and processes. Contractor attendance was very high at all of the tutorials, and they gave valuable input. The involvement of all of these organizations improves internal customer/supplier relationships, helps provide some focus on internal processes as well as external, and provides many additional people with direct experience working with TQM tools, and operating under the corresponding philosophy. All of these people will have to be supported by their management in carrying out these activities as they move into future activities.

Improving product and/or service quality is achieved through improvements in the processes that produce the product or service. Every activity and every job is part of a process - and can be improved. Improvement comes through people and learning. A strategy for process improvement used by teams helps provide a roadmap for further improvement. The roadmap includes the development of team members who possess the ability to determine a common objective, define the relevant process, define the current knowledge, and build on that knowledge to make a change in the process using the improvement cycle. If the enthusiasm, and team activities, initiated in the projects described in this paper are used as a roadmap into the future, the ACRV office and other organizations within NIO can realize significant improvements in processes, and can leave a well-defined trail for others to follow along the road of continuous improvement.

N93-26075

DISTRIBUTED PROJECT SCHEDULING AT NASA: REQUIREMENTS FOR MANUAL PROTOCOLS AND COMPUTER-BASED SUPPORT

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1992

Johnson Space Center

Prepared By:	Stephen F. Richards, Ph.D.
Academic Rank:	Associate Professor
University & Department:	Ambassador College Computer Information Systems Department Big Sandy, Texas 75755
NASA/JSC	
Directorate:	Information Systems
Division:	Software Technology
JSC Colleague:	Chris Culbert
Date Submitted:	August 7, 1992
Contract Number:	NGT-44-005-80

ABSTRACT

The increasing complexity of space operations and the inclusion of interorganizational and international groups in the planning and control of space missions lead to requirements for greater communication, coordination, and cooperation among mission schedulers. These schedulers must jointly allocate scarce shared resources among the various operational and mission oriented activities while adhering to all constraints. This scheduling environment is complicated by such factors as the presence of varying perspectives and conflicting objectives among the schedulers, the need for different schedulers to work in parallel, and limited communication among schedulers. Smooth interaction among schedulers requires the use of protocols that govern such issues as resource sharing, authority to update the schedule, and communication of updates. This paper addresses the development and characteristics of such protocols and their use in a distributed scheduling environment that incorporates computer-aided scheduling tools. An example problem is drawn from the domain of space shuttle mission planning.

INTRODUCTION

Scheduling is the process of assigning resources and times to each activity of a plan (or plans) while ensuring that each constraint is obeyed. Optimization criteria can determine the relative desirability of two alternate schedules. Although scheduling problems are often simple to visualize and express, scheduling is an NP-complete problem, so attempts to apply mathematical programming to scheduling have met with very limited success [2, 6]. In fact, programming approaches have been limited to very narrow problem domains, especially that of the job-shop, in which jobs must be assigned to various machines.

This paper focuses on the class of scheduling problem in which:

1. activities have precedence relationships (one activity must not begin until another activity has completed);
2. resources are limited;
3. objectives or optimization criteria exist that may be used to rank competing schedules; and
4. the time frame in which to complete all activities (or as many activities as possible) is limited.

This class of problem differs from the job-shop problem domain in that a job-shop problem assumes an infinite time line in which all activities may complete. In a job-shop problem, all activities are scheduled regardless of the total time required. In contrast, in this paper resources may be over-subscribed, so that even the optimum schedule might not accommodate all desired activities within the time limitations. Thus, provision must be made for selecting between competing activities (or sets of related activities) where insufficient time exists for the completion of all activities.

To assist users in developing viable schedules, NASA has developed COMPASS (COMPuter Aided Scheduling System) [3], a computer-based tool that interactively schedules activities in a user-specified order. COMPASS provides graphical tools for displaying activities, resource availability, and schedules. An activity defined in COMPASS may have precedence requirements and require resources. Activity attributes supported by COMPASS include priority, required resources, duration, earliest permissible start time, latest permissible end time, and state conditions. COMPASS has enjoyed widespread acceptance and use within NASA and the contractor community.

NASA has recently proposed enhancing COMPASS to support multi-user or distributed scheduling problems. This paper focuses on the issues raised by distributed scheduling and on requirements for computerized support of this problem domain. The next section of this paper defines distributed scheduling and addresses these issues. This is followed by a discussion of

some of the human issues involved in the development of protocols for use by multiple teams of schedulers who must cooperate to produce joint schedules.

DISTRIBUTED SCHEDULING

Definition

Distributed scheduling consists of those scheduling problems involving:

1. several schedulers,
2. who can work independently,
3. each of whom is responsible for scheduling separate sets of activities that are somehow interrelated, and
4. must share a common pool of resources.

Besides sharing a common resource pool, the activities may also have precedence requirements, or one activity may establish a state that another activity requires, etc. While the schedulers may work independently, the need to coordinate the interactions among their tasks prohibits purely independent work. Distributed scheduling problem domains of particular interest to NASA include the scheduling of astronomical satellite experiments, personnel training, and space mission activities.

The interactions among schedulers can be cooperative or competitive. Cooperative scheduling is defined as those cases in which:

1. All schedulers have the same objectives;
2. Responsibility for scheduling has been divided in order to share the labor; and
3. Protocols serve primarily to coordinate and synchronize.

In large problems the size and complexity of the scheduling task and the limited abilities, skills, knowledge, and resources of any individual make the distribution of the scheduling task a natural and necessary means of developing the required schedule. By distributing the work, each scheduler can concentrate on a manageable volume of work in a narrow domain. Some schedulers may develop specialized knowledge and skills suitable only to their particular domains.

In contrast, competitive scheduling consists of those cases in which:

1. Each scheduler has his or her own objectives;
2. The necessity of sharing common resources interferes with the simultaneous achievement of these objectives;
3. The pursuit of individual objectives leads to competition for the common resources; and
4. Protocols serve largely to arbitrate competition by allowing all schedulers fair access to

shared resources.

Competitive scheduling can arise in situations in which there are contractual agreements among different parties or in which different resources are owned by different groups. Such situations can dictate the division of responsibility for scheduling among several groups, with each group having its own set of goals.

General Discussion of Goals

The lack of a single point of control increases the complexity of the overall scheduling problem (as a result of the necessary communication overhead) and raises several issues regarding the interactions of multiple schedulers and the integration of their individual schedules. The most basic issue raised by distributed scheduling is that of goals. What measure of goodness is most appropriate in a distributed environment? How do the optimization criteria for a distributed scheduling problem differ from those for a non-distributed problem? Variables commonly used for scheduling problems include [4, 5]:

- Completion time: the time at which processing of the last activity completes.
- Flow-time: the total time that activities spend in the shop.
- Lateness: the difference between the completion time of an activity and some pre-specified due date associated with that activity.
- Tardiness: equal to lateness when lateness is positive, otherwise equal to zero.

Schedule evaluation criteria typically involve minimizing or maximizing the mean, total, minimum, or maximum of one or more of these variables. In a standard job-shop problem, these criteria are assumed to be universally agreed upon. However, even in such a standard, non-distributed scheduling environment, the various tasks to be scheduled may belong to several different customers (perhaps represented by members of the marketing staff), each of whom would prefer that his or her tasks be given high priority. Thus, even in a non-distributed setting conflicting goals may exist. When conflict exists, the scheduler must have some means of determining a set of priorities to be applied to the scheduling task. The scheduler may be flexible in his or her choice of priorities, adjusting them to the needs of the moment. For example, the scheduler might attempt to mollify a major customer who has previously been slighted by giving preference to that customer's work in the current schedule. Regardless of the conflicting demands, however, the optimization requirements are formulated under a single point of control and this procedure can succeed because the single scheduler (or team of schedulers) who develops the optimization criteria also controls the entire resource pool.

In a distributed schedule, however, individual schedulers must share resources, so one scheduler optimizing his or her schedule may restrict another scheduler's options, resulting in a subopti-

mal global schedule. The issue of a global measure of goodness becomes more important in distributed scheduling than in individual scheduling. This is true because an individual scheduler can accept a schedule even without a specific measure of goodness; the schedule may balance several conflicting needs fairly and "just look good." A distributed schedule, in contrast, must "look good" through several sets of eyes. When a team of schedulers must continue to work together on future projects, perceptions of inequity or misplaced priorities can engender resentments that will poison these on-going relationships. Thus, some mechanism for balancing both local and global optimization must be provided. The protocol used by the schedulers to coordinate their activities must support optimization techniques that are perceived as both equitable and efficient.

Requirements for Competitive Scheduling

NASA needs to develop protocols that facilitate the development of successful schedules in "competitive" distributed environments that generally satisfy the objectives of the separate schedulers. This requires protocols that govern the process of building the schedule as well as protocols that govern how conflicting objectives are resolved. Selecting a desirable scheduling protocol requires balancing several possibly conflicting requirements, including the following [1]:

1. The protocol should encourage the development of high quality schedules that score well when evaluated by either the global optimization criteria or the optimization criteria of individual schedulers. Where conflicting objectives exist, the protocol should lead to a reasonable compromise.
2. The protocol should be easy to understand, use, and implement. Features enhancing ease of use include ease of learning; minimum complexity; informative to the user of the state of activities, resources, etc.; and natural representation of concepts. Yet the process should be sufficiently rich in features and notation to encompass a wide range of scheduling problems.
3. The protocol should be mechanical and unambiguous.
4. The protocol should be general enough to work with a wide range of scheduling software. Schedulers should not be constrained to use a particular scheduling system or even the same system.
5. The resulting schedules should be resilient to unexpected changes.
6. The overhead should be kept to a minimum. For example, the volume and frequency of communications should be low.
7. The time required to develop schedules should be short, especially in highly dynamic environments.

8. Any computerized support should have a short response time. This requires that optimization techniques be computationally simple.
9. Rescheduling (the repair of a schedule because of unexpected occurrences, such as delays and loss of resources) must be especially fast.

Several sample scheduling techniques are listed below. The alternatives are discussed in terms of division of resources, communication, cooperation, and optimality.

1. Schedule tasks by priority. This approach requires that all tasks be known and prioritized in advance and then be scheduled in priority sequence. This is really non-distributed scheduling, except that we have several schedulers responsible for collecting tasks and we may provide improved computer support to enable the individual schedulers to track their own set of tasks by viewing only their portion of the schedule. This protocol also requires some mechanism for assigning priorities to tasks, such as a central authority or a voting scheme. (Schedulers with conflicting objectives may never agree on the assignment of priorities.) Although participants may perceive this method as fair (since no lower priority activity will be scheduled while a higher priority activity remains unscheduled), following this method strictly does not allow for compromises, such as scheduling two medium priority, low resource intensive activities instead of one higher priority, high resource intensive activity.
2. First come, first served. In this approach all schedulers are equal and none has priority over the others. Resources are not assigned to individual schedulers, but may be reserved by any scheduler. No cooperation among schedulers is required. Optimization is poor, because no attempt is made to balance the needs of multiple schedulers. There is a tendency among schedulers to reserve resources early, even before they know their full requirements. This hoarding can result in the allocation of resources to low priority tasks.
3. Divide resources among schedulers in advance. This method permanently allocates resources to specific schedulers who can use them as they choose. No communication or cooperation among schedulers is required. Schedulers need not even know the global schedule. This approach is impractical when there is a potential state conflict between tasks (e.g., when two schedulers independently schedule a treadmill experiment and a microgravity experiment that requires no vibration). This approach may also yield poor schedules when one scheduler assigns resources to low priority tasks or leaves resources unused that could be used by another scheduler. In this approach the quality of the resulting schedule is limited by the appropriateness of the initial allocation of resources. A poor allocation may result in few activities being successfully scheduled.

4. Divide resources among schedulers in advance but permit borrowing. This approach differs from the previous one by permitting schedulers to negotiate among themselves to improve their schedules. There is still no need for global optimization criteria. The status and bargaining power of individual schedulers is determined by the initial allocation of resources. Communication needs consist of a knowledge of resources available to other schedulers.
5. Sharing of intentions among schedulers. In this approach schedulers review their intentions with their peers and receive feedback before reserving resources and committing to a particular schedule. While this approach has the potential for producing high quality schedules through the sharing of knowledge and expertise, it also imposes a heavy communication burden among schedulers that can negate much of the benefit resulting from distributing the scheduling task. This approach is also fragile in that its success depends on the voluntary cooperation of each scheduler. Where this cooperation fails, this approach can degenerate into a first come, first served system.
6. Simultaneous iterative scheduling. In this method each scheduler devises a schedule and shares it with others. Schedulers identify and resolve conflicts by some agreed upon method. If unscheduled tasks and unallocated resources remain, another round of scheduling follows. In this approach all schedulers must be ready to schedule simultaneously. Also, each participant must be provided some incentive to cooperate with the others in resolving conflicts. The global schedule must be available to all schedulers.
7. Consecutive iterative scheduling. In this method the schedulers are divided into two or more groups that alternately devise schedules. This approach is useful when one group creates resources required by another. For example, a university administration develops a schedule of classes, the students then submit their individual schedule requests, and the administration, after analyzing the requests, adds sections to some classes and deletes sections from others. The students then request changes to their schedules. In principle this cycle can continue for many iterations. This approach requires some incentive to cooperate and requires that each scheduler knows the global schedule and the state of available resources.

Any attempt to develop a universal scheduling methodology is doomed to failure because of the enormous diversity of scheduling domains. The variety of tasks, resources, constraints, and environments is virtually unlimited. The methodologies listed above are not applicable to all domains but must be selected based on the characteristics of the specific domain of interest.

Several other issues that are particularly relevant to distributed scheduling are briefly addressed in the remainder of this section. One of these is the requirement for revising a schedule,

also termed rescheduling [2]. Several factors can trigger a need to reschedule. A resource can become unavailable, making the current schedule unfeasible; a task can take longer than expected; or a user can change his or her requirements so as to impose a conflict, exhaust a resource needed by a later task, or delete an enabling task that creates a subsequently needed resource or state. In addition, rescheduling is desirable, although not required, whenever an opportunity arises to improve the schedule by adding previously unscheduled tasks or resequencing already scheduled tasks. This can happen, for example, when new resources become available or when a task completes early. Differences between scheduling and rescheduling include:

1. Rescheduling takes place in the context of an existing schedule that we may wish to disturb as little as possible;
2. Rescheduling must consider work in progress;
3. Rescheduling often must occur quickly, in contrast to the initial scheduling which may be performed in a more leisurely manner; and
4. Someone other than the original scheduler may perform the rescheduling.

An important issue for rescheduling in a distributed scheduling environment is the need to reduce communication requirements among schedulers to facilitate quick rescheduling. Since this may require a return to centralized scheduling, the rescheduler must have the appropriate information to make beneficial changes.

Another issue is that of database support for distributed scheduling. A distributed scheduling system requires many of the features of a distributed database management system. The system must merge separate databases of tasks, resources, constraints, and assignments into a single image while retaining the ability to display for individual schedulers only those portions of the database under their control. However, since each scheduler has a different view of the world (with different granularity levels, time scales, measures of goodness, types of constraints, etc.), the system must support different user languages and communicate with each scheduler in a natural and helpful way. As our software tools, such as COMPASS, address more diverse and complex problem domains, we will require a more comprehensive database language for describing scheduling problems.

Communication and coordination among schedulers in an important issue. NASA schedulers who impact one another may work at different centers, making communication difficult. The scheduling of Space Station Freedom will involve groups in several countries. An important research question concerns how frequently NASA schedulers communicate. Is the level of communication optimal? If it is below optimal, do schedulers fail to communicate because they do not perceive a need to communicate, or because they feel communication is too time consuming, or because they fear loss of control of their environment, or is there some other reason? If indepen-

dent scheduling is a human preferred approach, then it will be important to determine why this is true, how we can encourage people to cooperate, and how we can enhance cooperation while minimizing communication. The mechanisms for communication and coordination (the languages, database support, and interaction procedures) appear to be a critical aspect of distributed scheduling by human agents.

A final issue involves the introduction of expert system support for scheduling. Optimization heuristics have been envisioned for individual scheduling support; some of this support is already available on COMPASS. Expert system support for distributed scheduling would focus on communication and negotiation. An expert system that monitored the actions of all schedulers could infer when one scheduler needed to know of the actions of another. This technique could reduce communications requirements among human schedulers. Also, an expert system could search for instances in which two schedulers could trade resources or reschedule certain activities to their mutual advantage. Ultimately, we may wish to introduce artificially intelligent schedulers into a distributed scheduling system. The scheduling of certain domains, such as power generation, may be suitable for AI approaches. Once AI schedulers are developed for individual scheduling domains the natural next step would be to introduce them into human scheduling systems. This possibility raises questions regarding how artificial and human schedulers might best interact.

SAMPLE PROTOCOL: THE RED-BLUE PROTOCOL

The Red-Blue protocol has been devised to guide the interactions between schedulers at NASA/JSC and SpaceHab, Inc. of Huntsville, Alabama as they schedule STS-57, due to launch in April, 1993. The objective of this protocol is to facilitate the production of payload deployment and management schedules in the context of other orbiter/station operations. Based on our experience with scheduling this mission, the Red-Blue protocol will be enhanced and used as NASA's standard protocol for the distributed scheduling of shuttle (and, later, space station) operations.

The requirements for the Red-Blue protocol are similar to those discussed above. It should be easy to understand, use, and implement. Its use should be mechanical and unambiguous. It should work with a variety of scheduling software systems. It should support rescheduling. Its requirements for communication among schedulers, in terms of frequency and volume of communication, should be low. It should allow the creation of schedules in a timely manner. Finally, where there are conflicting objectives, the protocol should lead to the creation of schedules that provide a reasonable compromise between these objectives.

The Red-Blue protocol begins by dividing all of the activities into two groups, red and blue. The red activities can only be scheduled by the red scheduler, in this case, NASA. Likewise, the blue activities can only be scheduled by the blue scheduler, SpaceHab, Inc. Limits can be placed

on the volume of resources that can be used by the red and blue schedulers; for example, a scheduler might be limited to a maximum quantity of water during the mission or a maximum number of hours of an astronaut's time. Limits are not placed, however, on where in the schedule the resources may be used. In the case of NASA and SpaceHab, Inc., these limits have been established during contract negotiations. Any subsequent modifications or clarifications to these limits must be worked out by the schedulers and possibly their management.

The red scheduler produces the first schedule, placing red activities anywhere on the timeline, up to the limit of the red resource allocation. For example, the red scheduler would schedule basic activities such as course correction burns and astronaut sleep and meal times as well as mission specific activities such as payload deployment. Next the blue scheduler places blue activities in any available (white) space on the timeline, up to the limit of the blue resource allocation. Thereafter, only one scheduler may work on the timeline at a time. The scheduler who has authority to modify the schedule at any particular time is said to "hold the token." When the other scheduler has activities to schedule, that scheduler may request the token. The scheduler who holds the token may schedule or move his or her activities within any white space and within any space that he or she already occupies. The scheduler may not, however, move any activities of the other scheduler, or oversubscribe any resource.

If one scheduler wants to move an activity into the space owned by the other scheduler, the two schedulers can negotiate a set of changes that can then be produced by operations according to the basic protocol. While this protocol assumes that the parties are competitive (having differing and possibly conflicting goals), it also assumes that they are not antagonistic. Thus, the protocol assumes that the parties will cooperate whenever the result of such cooperation leaves neither party worse off.

A low communication procedure for asking the other party to move some of its activities is to allow a scheduler to oversubscribe resources (thereby producing a conflict between red and blue activities). The other scheduler, when he or she next holds the token, can leave the oversubscription (thereby delaying the resolution of the conflict), unschedule the offending activities of the other color, or accommodate his or her counterpart by moving some activities of his or her own color.

If more than two schedulers need to work cooperatively, then the Red-Blue protocol can be extended by devising a procedure for exchanging authority to operate on the schedule. A research question is to investigate the social and communications changes that occur as the number of scheduling groups rises.

Several implementation issues must yet be addressed. When activities must be rescheduled during a mission, does one party have the right to force the other to modify its schedule? For

example, if an activity runs long, can the other scheduler force termination of the activity? Also, electronic protocols must be developed for the exchange of schedule updates. These protocols must enforce the requirement that only the token-holder may modify the schedule and must ensure that all parties always agree on the composition of the current schedule. Thus, when one party refers to the current schedule or to the schedule as it existed two versions ago, the other party will know what is meant.

CONCLUSION

NASA has an unlimited variety of distributed scheduling problems. Competitive distributed scheduling problems arise when shared use of common resources interferes with the simultaneous achievement of multiple resources. We need to develop protocols that govern the process of building the schedule and govern how conflicting objectives are resolved. These protocols can only be developed and evaluated in the context of specific applications. This paper presents a simple, yet effective Red-Blue protocol to facilitate the production of payload schedules in the context of other orbiter/station operations.

REFERENCES

[1] Fox, Mark S., "Constraint-Directed Search," Ph.D. dissertation, Carnegie-Mellon University, 1983.

[2] Fox, Mark S. and Zweben, Monte, "Knowledge Based Scheduling," Tuturial MA2, presented at the Ninth National Conference on Artificial Intelligence, July 15, 1991.

[3] McDonnell Douglas Space Systems Co. "COMPASS 2.0 User's Manual," for NASA/Johnson Space Center Software Technology Branch, 1991.

[4] Ow, Peng Si, "Heuristic Knowledge and Search for Scheduling," Ph.D. dissertation, Carnegie-Mellon University, 1984.

[5] Salvador, Michael S., "Scheduling and Sequencing," in *Handbook of Operations Research*, Joseph J. Moder and Salah E. Elmaghraby (Eds.), Van Nostrand Reinhold, New York, 1978, pp. 268-300.

[6] Ullman, J. D., "NP-Complete Scheduling Problems," in *Journal of Computer and System Sciences*, Vol. 10, 1975, pp. 384-393.

N93-26076

DEVELOPMENT OF A PYROLYSIS WASTE RECOVERY MODEL WITH DESIGNS, TEST PLANS, AND APPLICATIONS FOR SPACE-BASED HABITATS

Final Report

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM--1992

Johnson Space Center

Prepared By:	Bobby J. Roberson, Ph.D.
Academic Rank:	Professor
College and Department:	Bevill State Community College, Division of Natural Sciences, Fayette, Alabama 35555
NASA/JSC	
Directorate:	Engineering
Division:	Crew and Thermal Systems
Branch:	Life Support Systems
JSC Colleague:	Charles E. Verostko
Date Submitted:	August 7, 1992
Contract Number:	NGT-44-005-803

ABSTRACT

Extensive literature searches revealed the numerous advantages of using pyrolysis as a means of recovering useable resources from inedible plant biomass, paper, plastics, other polymers, and human waste. A possible design of a pyrolysis reactor with test plans and applications for use on a space-based habitat are proposed. The proposed system will accommodate the wastes generated by a four-person crew while requiring solar energy as the only power source. Waste materials will be collected and stored during the 15-day lunar darkness periods. Resource recovery will occur during the daylight periods. Useable gases such as methane and hydrogen and a solid char will be produced while reducing the mass and volume of the waste to almost infinitely small levels. The system will be operated economically, safely, and in a non-polluting manner.

INTRODUCTION

In a controlled ecological life support system (CELSS) such as a space-based (lunar/Mars) habitat, water, food, and oxygen must be stored and recycled. The problems of mass and volume limitations and infrequent resupply dictate that artificial methods be employed to supply the needs of the inhabitants (15). One of the most important aspects of a CELSS involves a closed loop system (14). A CELSS must have the ability to: a) process human waste and plant biomass and recover resources from these materials; b) recondition and revitalize the air by using waste materials; c) maintain a mass balance within the system; d) provide safe and reliable life support for long-duration missions; e) attain a high level of self-sufficiency; f) lower the expense and simplify the problems related to resupply; and g) provide an Earth-like environment that is conducive to productivity (6,11).

Requirements of a CELSS

The waste resource recovery portion of a CELSS will consist of a waste and water management subsystem, an atmosphere management subsystem, and a food management subsystem (5). The outputs of one subsystem should match the requirements of the others while observing quality and quantity tolerances (6). The design of such a CELSS should also permit future expansion, and should be safe, reliable, flexible, and easily maintained. A CELSS that incorporates biological and physicochemical (B/PC) technologies to recycle water, air, food, and biomass holds the ultimate promise of allowing humans to become successfully established outside the Earth's biosphere (10).

Spaceflights through the Space Shuttle era have not been dependent on recycling and regeneration of useful resources because of the short duration of the flights. Instead, waste materials have been stabilized, stored, and returned to Earth for analysis and disposal. This will not be feasible on lunar and Mars missions. The majority of the food on future, long-duration flights and missions must be grown in a CELSS. Most of the plant biomass is inedible and must be processed to supply carbon dioxide, water, and nutrients for plant growth, or fuel for other systems. It is estimated that inedible plant biomass and paper products will constitute approximately half of the total dry, solid waste material generated in a CELSS (13).

The simplest, and probably the most obvious, method of recovering some of the value from any organic material is by burning it for its energy

content (4). The two processes used for the destruction of materials by heating are incineration and pyrolysis.

Incineration

Incineration is the complete oxidation of waste using either pure oxygen or oxygen diluted with an inert gas. A temperature of 1000°C and a pressure of approximately one atmosphere are required. All types of combustible materials such as inedible plant biomass, paper, plastics, food waste, human waste, etc. may be treated. The incineration process usually involves two major steps: a) initial heating to sterilize and to vaporize the volatiles; and b) combustion of the dry waste. The process is relatively fast and results in 97 to 99% reduction of the waste. The end product is usually a fine, dry powder (7). Most incinerators contain two chambers. In the primary chamber auxiliary burners dry, volatize, and ignite the waste. The amount of air present is in 75 to 100% excess. When combustion is sustained, the burners are turned off. The gases and combustibles flow into the secondary chamber for burn completion. The air is maintained at an excess level of 75 to 150%. The excess air, turbulence, and retention time provide conditions for complete combustion (12). The chemical reaction for the incineration of cellulosic materials is:

$$C_6H_{10}O_5 + 6\,O_2 + \text{heat} \dashrightarrow 6\,CO_2 + 5\,H_2O$$

Pyrolysis

Pyrolysis is the chemical destruction of a carbonaceous material by heating to 400 to 1000°C in the absence of oxygen, or in a controlled oxygen environment. The pyrolytic process results in the formation of four phases of products: a) a gaseous component including hydrogen, methane, carbon monoxide, and carbon dioxide; b) an oil including organic acids, alcohols, ketones, etc.; c) a char including carbon and inert compounds; and d) an aqueous phase containing some water-soluble compounds. In general, the faster the decomposition of the materials, the higher the yield of gas, whereas a long, slow heating process results in higher proportions of oil and char. Also, upon rapid heating in the absence of oxygen, cellulose molecules "explode," and the fragments are free to form simple compounds such as methane, carbon monoxide, hydrogen, and water (1). During the pyrolytic process, cellulose increases in porosity and swells as volatiles are evolved. Pyrolysis begins at approximately 200°C,

and tars evolve. As the temperature rises the products decompose or crack, forming hydrogen-rich gaseous compounds and solid carbon which approaches graphitic carbon. The pyrolysis of cellulose reaches completion between 600 and 800°C (2). The overall reaction for the pyrolysis of cellulose is:

$$C_6H_{10}O_5 + \text{heat} \rightarrow CH_4 + 2\,CO + 3\,H_2O + 3\,C$$

The reaction appears to be endothermic at lower temperatures and exothermic at higher temperatures. Retention time is usually between 12 and 15 minutes (8). Some of the gases produced are combustible and have value by providing heat to the effluent gases. The carbon residue, a char, can also be used as a fuel. No oxygen is added, and the starting material is only cellulose. Since the original material is not necessarily pure cellulose, the effluent gases may contain some simple and complex organic compounds. The char may also contain some minerals, ash, and other inorganics as well (4).

The major concerns of a pyrolysis and/or incineration system in a space-based habitat are the use of high temperature, explosion potential, energy requirements, fire propagation, production of toxics, required maintenance, efficiency, mass and volume requirements, pre- and post-treatment of waste, requirement for a continuous system, and longevity of the system (13).

DISCUSSION

Comparison of the advantages and disadvantages of pyrolysis versus incineration indicate that pyrolysis might possibly be the most appropriate physicochemical combustion technology for the recovery of resources from waste materials. The advantages of pyrolysis include the following: a) greater stability with little response to change in feed rate and therefore easier to control; b) 15 to 25% higher process rates (feed capacity per square foot of bed area); c) fewer and smaller particulates; d) less or no supplemental fuel required; e) generation of useable gases; and f) smaller mass and volume waste requirement (3). A pyrolysis reactor on the lunar surface will require no fuel other than that supplied by the sun through the use of solar concentrators. Also, a purge gas may not be needed since the vacuum of space could be used to evacuate the air from the chamber. In some instances a combined incineration/pyrolysis system might be employed.

The actual results of a commercial pyrolytic process are shown in Table 1. The effluent gases of three wastes indicate high percentages of two gases that are potential fuels (methane and hydrogen).

TABLE 1.- PYROLYSIS TEST RESULTS*

	Type of Waste		
Gases Produced (mol %)	***Liquid Organic***	***Garbage***	***Drugs***
Hydrogen	34.60	30.29	---
Carbon dioxide	6.22	14.19	---
Ethylene	8.52	8.06	10.37
Ethane	2.35	0.91	4.13
Acetylene	0.15	0.43	0.01
Oxygen	---	---	---
Nitrogen	0.70	0.48	---
Methane	25.53	10.80	78.78
Carbon monoxide	6.22	26.25	---
Propane "plus"	15.71	8.58	6.71
Hydrogen sulfide	---	---	---

* Test results are from a pyrolysis system invented, owned, and operated by Benjamin P. Fowler, Fowler Engineering Company, Houston, Texas. Complete descriptions are given in U.S. Patent No. 4,934,286. The mention of a vendor does not imply endorsement by NASA.

Proposed System

The proposed system will operate at high temperatures (up to 1000°C). The power source for the initial test model will be either an electrical furnace supplying heat to ceramic plates, a carbon arc, a graphite resistance furnace, or a gas burner. The actual space-based system will use solar concentrators, or electrical power from solar panels or a nuclear power source. The system will accommodate plant biomass, plastics, other polymers, human wastes, etc.

Figure 1 is a schematic of the proposed pyrolytic process.

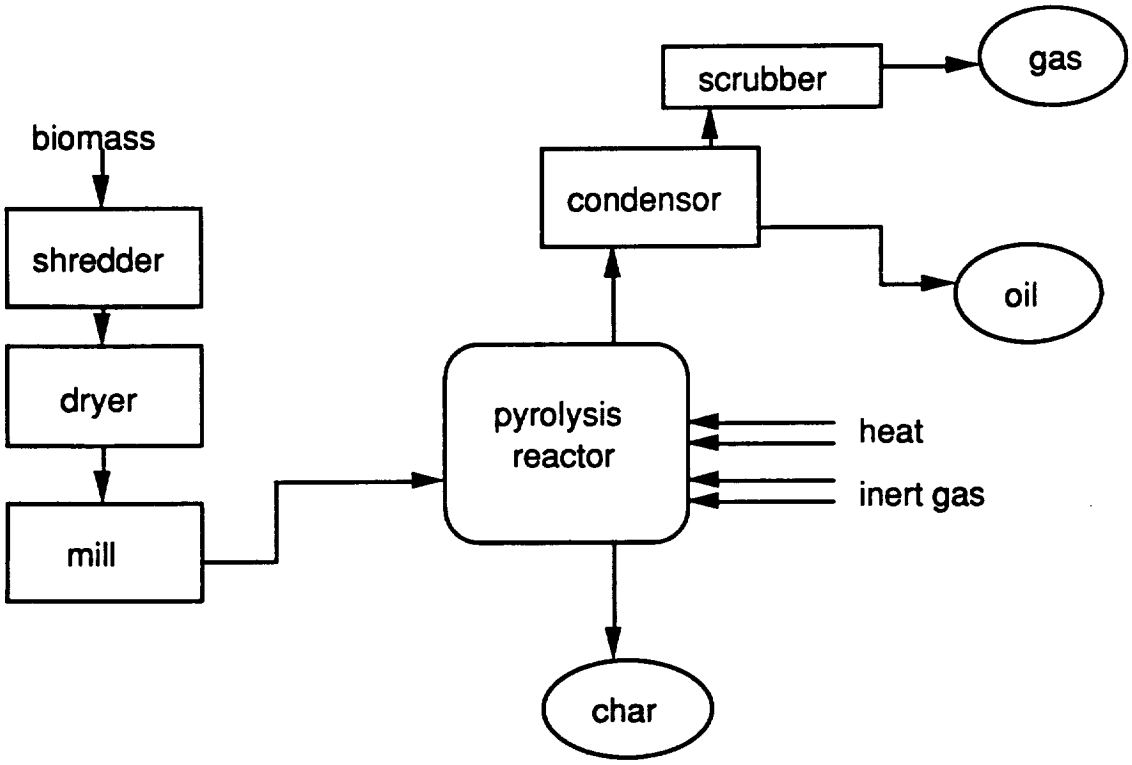


Figure 1.- Schematic of the Pyrolysis Process

Proposed Test Plans

The pyrolysis reactor shall be tested using several waste materials, individually and collectively. The test materials should include inedible plant biomass, paper, plastics, other polymers, and possibly human biomass waste. Test conditions shall be varied to attain the optimum rate and degree of pyrolysis with the recovery of all products for analysis. Useable products should be identified for each type of waste. The test conditions shall initially include the items shown in Table 2.

TABLE 2.- PYROLYSIS TEST CONDITIONS

Maximum temperature	1000°C
Pressure	ambient
Amount of biomass	10 kg
Moisture content	< 3%

Steps in the Pyrolysis Process

The steps in the pyrolytic process are outlined in Table 3.

TABLE 3.- PYROLYSIS PROCEDURES

1. Biomass is collected.
2. Biomass is shredded.
3. Biomass is dryed to reduce moisture content to 3% or less.
4. Biomass is milled to reduce particle size to 30 microns or less.
5. Biomass is fed into the pyrolysis reactor vessel.
6. Oxygen and air are expelled from the reactor by the introduction of an inert gas or by evacuation to the vacuum of space.
7. The reactor vessel is sealed to prevent the introduction of additional oxygen or air.
8. The vessel is heated to 1000°C where pyrolysis is carried to completion while gases are collected.
9. Effluent gases are scrubbed, separated, and condensed if necessary.
10. The system is shut down for maintenance if required.

Design of the Pyrolysis Unit

The pyrolysis unit will be constructed, tested, and monitored under experimental conditions. The prototype laboratory model should be built to the specifications of a lunar/Mars model and should be tested under the

conditions that exist on the surfaces of those bodies. Careful monitoring of effluent gases should be conducted to assure that no toxic gases will be released to space or to the atmosphere of Mars. Analysis of the products of the pyrolysis of various substances should also be conducted to determine their uses as fuels, nutrients, etc. Numerous pyrolysis reactors in commerical use offer suggestions for a possible space-based model. Some of the criteria for a lunar/Mars pyrolysis reactor system are listed in Table 4.

TABLE 4.- PYROLYSIS REACTOR SPECIFICATIONS

1. The reactor vessel shall be relatively compact.
2. The reactor vessel and all hardware shall be constructed of corrosion proof materials.
3. The system shall tolerate temperatures up to 1000°C.
4. The system shall accommodate a minimum of 10 kg of waste per day.
5. The system shall operate in a batch mode.
6. The system shall permit the recovery of useful gases and and solids.
7. The system shall be easily maintained.
8. The system shall be safe and reliable.

Figure 2 shows a possible design configuration for a pyrolysis reactor.

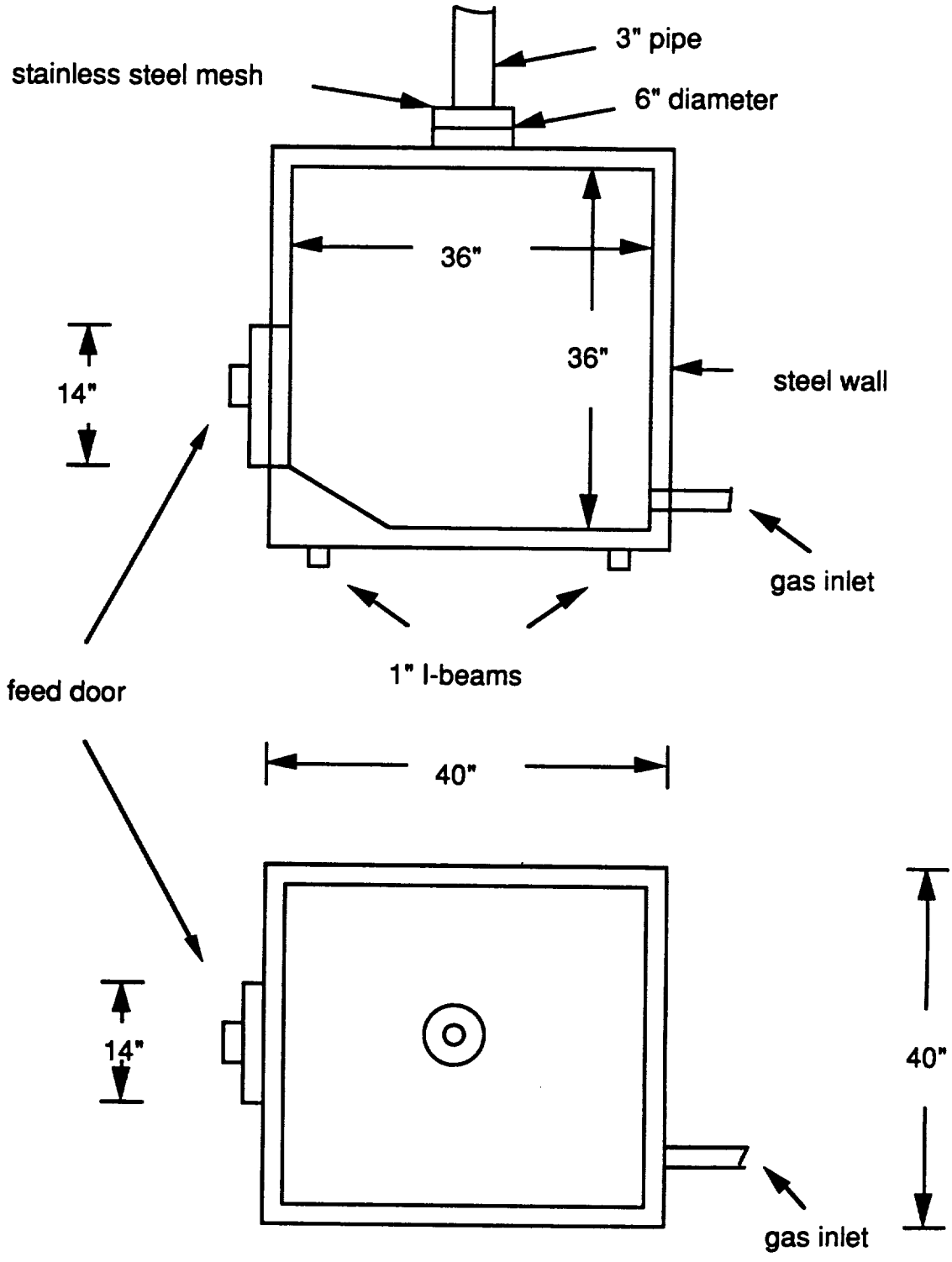


Figure 2.- Pyrolysis Model

CONCLUSIONS

Pyrolysis appears to be a highly efficient method for the recovery of useful resources from inedible plant biomass, paper, plastics, other polymers, and even human waste. Solar furnaces can supply the power for the operation of the unit while methane and hydrogen can be collected for use as fuels. The hydrogen could also be used to produce water (assuming that oxygen can be obtained from the minerals on the moon and Mars). The gases can be scrubbed to remove possible toxic materials. The char and oils resulting from the pyrolysis process can also be used as fuels. The system will be capable of batch mode operation and can possibly be integrated with other physicochemical and biological systems of a CELSS. The system should be relatively inexpensive, easily maintained, and should be extremely durable and safe. Further studies are essential including the actual construction and testing of the pyrolysis unit.

REFERENCES

1. Barton, A. F. M. *Resource Recovery and Recycling.* 1979. John Wiley and Sons. Pp. 418-425.

2. *Combustion and Incineration Processes.* W. R. Niessen, Ed. 1978. Marcel Dekker. Pp. 68-75.

3. *Design of Sewage Sludge Incineration Systems.* C. J. Brunner, Ed. 1980. Noyes Data Corp.

4. *Handbook of Incineration Sytems.* C. R. Brunner, Ed. 1991. McGraw-Hill.

5. Henninger, D. L. "Life Support Systems Research of the Johnson Space Center" in *Lunar Base Agriculture: Soils for Plant Growth.* D. W. Ming and D. L. Henninger, Eds. 1989. American Society of Agronomy. Pp. 173-192.

6. Hoshizaki, T. and B. D. Hansen, III. "Generic Waste Management Requirements for a Controlled Ecological Life Support System (CELSS)." NASA Conference Publication 2247. May 3-6, 1981. Pp. 38-49.

7. "Housekeeping Concepts for Manned Space Flights." Final Report, Manned Space Systems Advanced Programs, Fairchild Hiller, Republic Aviation Division. October 30, 1979. Pp. 4/39-4/40.

8. *Incineration Systems.* C. R. Brunner, Ed. 1984. Van Nostrand-Reinhold Co.

9. *Pyrolysis of Polymers, Vol. 13: Fire and Flammable Series.* C. J. Hilado, Ed. 1976. Technonic Publ. Co., Inc.

10. Rummel, J. and T. Volk. "A Modular BLSS Simulation System" in NASA Conference Publication. 2480. 1987. Pp. 57-66.

11. Schwartzkopf, S., J. Houle, and A. Guastaferro. "Lunar Base Controlled Ecological Life Support System (LCELSS)." Preliminary Conceptual Design Study. Final Report. April 3-10, 1991. Lockheed Missiles and Space Company, Inc.

12. *Small-Scale Resource Recovery Systems.* A. E. Martin, Ed. 1982. Noyes Data Corp. Pp. 235-329.

13. Verostko, C. E., N. J. C. Packham, and D. L. Henninger. "Final Report on NASA Workshop on Resource Recovery from Wastes Generated in Lunar/Mars Controlled Ecological Life Support Systems (CELSS)." May 1992. JSC Document No. 25736, CTSD-ADV-035.

14. Volk, T. and J. P. Rummel. "The Case for Cellulose Production on Mars" in *The Case for Mars III: Strategies for Exploration.* C. E. Stoker, Ed. 1989. Pp. 87-93.

15. Wyedeven, T., J. Tremor, C. Koo, and R. Jacques. "Sources and Processing of CELSS Wastes." Adv. Space Research 9: 85-97 (1989).

N93-26077

UTILIZATION OF THE GRADED UNIVERSAL TESTING SYSTEM TO INCREASE THE EFFICIENCY FOR ASSESSING AEROBIC AND ANAEROBIC CAPACITY

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1992

Johnson Space Center

Prepared By:	Sandra L. Rodgers, Ph.D.
Academic Rank:	Associate Professor
University & Department:	Louisiana Tech University Dept. of Health and Physical Education Ruston, LA 71272
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Medical Life Sciences
Branch:	Space Biomedical Research Institute
JSC Colleague	Steven F. Siconolfi, PhD.
Date Submitted:	August 27, 1992
Contract Number:	NGT-44-005-803

ABSTRACT

The in-flight exercise test performed by cosmonauts as part of the Russian Exercise Countermeasure Program is limited to 5 minutes due to communication restrictions. During a recent graded exercise test on a US Shuttle flight, the test was terminated early due to an upcoming loss of signal (LOS) with the ground. This exercise test was a traditional test where the subject's exercise capacity dictates the length of the test. For example, one crewmember may take 15 minutes to complete the test, while another may take 18 minutes. The traditional exercise test limits the flight schedulers to large blocks of space flight time in order to provide medical and research personnel information on the fitness capacity (maximal oxygen uptake: VO2max) of crewmembers during flight. A graded exercise test that would take a finite amount of time and a set preparation and recovery time would ease this problem by allowing flight schedulers to plan exercise tests in advance of LOS. The Graded Universal Testing System (GUTS) was designed to meet this goal.

Fitness testing of astronauts before and after flight provides pertinent data on many variables. The Detailed Supplemental Objective (DSO608) protocol (6) is one of the graded exercise tests (GXT) currently used in astronaut testing before and after flight. Test times for this protocol have lasted from 11 to 18 minutes. Anaerobic capacity is an important variable that is currently not being evaluated before and after flight. Recent reports (1,2,5) from the literature have suggested that the oxygen deficit at supramaximal exercise is a measure of anaerobic capacity. We postulated that the oxygen deficit at maximal exercise would be an indication of anaerobic capacity. If this postulate can be accepted, then the efficiency of acquiring data from a graded exercise test would increase at least twofold. To examine this hypothesis anaerobic capacity was measured using a modified treadmill test (3,4) designed to exhaust the anaerobic systems in approximately 45 to 75 seconds. Lactate concentration in the blood was analyzed after all tests, since lactate is the end-product of anaerobic energy production. Therefore, the peak lactate response is an additional indication of anaerobic capacity.

A preliminary comparison of the GUTS and the DSO608 suggests that the GUTS protocol would increase the efficiency of VO2max testing of astronauts before and after flight. Results for anaerobic capacity have not been tabulated.

INTRODUCTION

The in-flight exercise test performed by cosmonauts as part of the Russian Exercise Countermeasure Program is limited to 5 minutes due to communication restrictions. During a recent graded exercise test on a US Shuttle flight, the test was terminated early due to an upcoming loss of signal (LOS) with the ground. This exercise test was a traditional test where the subject's exercise capacity dictates the length of the test. For example, one crewmember may take 15 minutes to complete the test, while another may take 18 minutes. The traditional exercise test limits the flight schedulers to large blocks of space flight time in order to provide medical and research personnel information on the fitness capacity (maximal oxygen uptake: VO2max) of crewmembers during flight. A graded exercise test that would take a finite amount of time and a set preparation and recovery time would ease this problem by allowing flight schedulers to plan exercise tests in advance of LOS. The Graded Universal Testing System (GUTS) was designed to meet this goal.

Fitness testing of astronauts before and after flight provides pertinent data on many variables. The Detailed Supplemental Objective (DSO608) protocol is one of the graded exercise tests (GXT) currently used in astronaut testing before and after flight (6). Test times for this protocol have lasted from 11 to 18 minutes. Anaerobic capacity is an important variable that is currently not being evaluated before and after flight. Recent reports (1,2,5) from the literature have suggested that the oxygen deficit at supramaximal exercise is a measure of anaerobic capacity. We postulated that the oxygen deficit at maximal exercise would be an indication of anaerobic capacity. If this postulate can be accepted, then the efficiency of acquiring data from a graded exercise test would increase at least twofold. To examine this hypothesis anaerobic capacity was measured using a modified treadmill test (3,4) designed to exhaust the anaerobic systems in approximately 45 to 75 seconds. Lactate concentration in the blood was analyzed after all tests, since lactate is the end-product of anaerobic energy production. Therefore, the peak lactate response is an additional indication of anaerobic capacity.

The purpose of this study was to compare the relative abilities of the GUTS and DSO608 GXT protocols to assess VO2max. This study was also concerned with the feasibility of the DSO608 and GUTS protocols to obtain values of aerobic and anaerobic capacity and thereby increase the efficiency of acquiring data from a graded exercise test.

METHODS

Fifteen subjects participated in the study. All subjects underwent cardiovascular screening prior to testing utilizing the Bruce GXT protocol. A subject profile is presented in Table 1. Two GXT's were administered to measure maximal oxygen uptake (VO2max), and an anaerobic test was utilized to assess anaerobic capacity. Small blood samples were drawn to measure the lactate concentration before and after all tests. Blood lactate is an indicator of anaerobic responses.

The GUTS Protocol is a maximal treadmill test designed to assess VO2max in 12 minutes by setting speed and grade for each stage of the test based on the subject's estimated VO2max. The starting treadmill speed was based on 70% of the subject's age, sex, weight predicted VO2max. Subsequent increases are determined by the subject's heart rate response at the end of each stage. The test consists of six stages increasing in difficulty every three minutes through stage three and then every minute until the test ends at 12 minutes.

The DSO608 is the standard treadmill GXT (6) utilized to evaluate the VO2max of astronauts before and after flight. Speed increases from 3.5 miles per hour (mph) in stage one to 7 mph in stage four. Subsequent workload increases are accomplished by a grade change of 3% until the subject reports volitional fatigue.

An anaerobic test was administered in order to determine if selected measures obtained during aerobic GXT's reflect anaerobic capacity, and thereby eliminate the need for astronauts to perform an anaerobic test. The instrument was a modified (3) version of a test developed by Schnabel and Kindermann (4). Subjects were required to run at 8 mph, 20% grade until volitional exhaustion after a one minute warm-up run at 5 mph. To encourage subjects to push to reach true exhaustion, all subjects were placed in a modified parachute harness that would prevent them from falling on the treadmill.

Anaerobic capacity may be expressed as total work (kgm, kcal, or kj) performed during the treadmill run or as average power (6). The values are based on the product of the subject's body weight and total vertical distance covered during the test.

Total vertical distance (TVD) can be determined from test time utilizing one of the equations below:

Test time from 0 to 39 seconds:

$$TVD(m) = -5.304 \times 10^{-3} - (1.7391 \times 10^{-2} \times time) + (9.1581 \times 10^{-3} \times time^2)$$

Test time for 40+ seconds:

$$TVD(m) = 13.902 + (0.70079 \times (40\text{-}time))$$

Total work can be computed from the following:

$$Work\ (kgm) = Body\ Weight(kg) \times TVD(m)$$

$$Work\ (kcal) = Work(kgm) \times 2.3427 \times 10^{-3}\ \ (kgm/kcal)$$

$$Work\ (kJ) = Work(kcal) \times 4.186\ (kJ/kcal)$$

Average Power can be computed from:

$$Power(watts) = (Work(kgm) \times time(sec)/60)\ /\ 6.118$$

RESULTS/DISCUSSION

The DSO608 protocol is a GXT designed to assess VO2max by allowing the subject to exercise until exhaustion. The end of the test is determined by the subject and test times vary from 11 to 18 minutes. The GUTS test is designed to elicit VO2max in 12 minutes. Estimated and measured VO2max values were compared on both protocols to determine if a true VO2max was achieved or if the test measured VO2peak. On both tests the final two work rates were examined and the per cent change in work rate was compared to the per cent change in VO2. If the change in work rate was larger than the change in VO2 by 5% the VO2 was considered a maximum rather than a peak value.

Preliminary comparisons of the results would indicate that the GUTS protocol is a valid test for assessing VO2max (Figure 1). Validity is based on acceptable correlations, no physiologically significant differences and low standard error of estimates (SEE). The correlation for VO2max between the protocols was r= 0.85 with a SEE of 0.40 L/min or ~10% of the DSO608 mean value. The mean difference between the measures was 0.25 L/min or ~6%. This difference was statistically significant (p=0.0346) but is close to being

within measurement error (5%). The final external work rates for the two protocols were not significantly (p=0.4467) different. The final external work rate for the DSO608 protocol was 327 (±72) watts while the final work rate achieved with the GUTS protocol 321 (±64) watts.

Two individuals were unable to complete the GUTS protocol due to very high work rates (higher than that achieved on DSO608). An analysis of the data without these individuals improved the correlation to an r of 0.89 with a mean difference of 0.07 L/min. This suggests that the basic approach of the GUTS protocol is valid for 80-90% of the subjects tested. An improvement in the protocol may be accomplished by utilizing other submaximal data in the setting of the external work rates.

The relative abilities of the DSO608 and GUTS protocols to achieve true maximal work were examined. A physiological maximum (for VO2) was reached 80% of the time for DSO608 and 40% of the time for GUTS. Only 47% of the people reaching max on DSO608 reached max on GUTS.

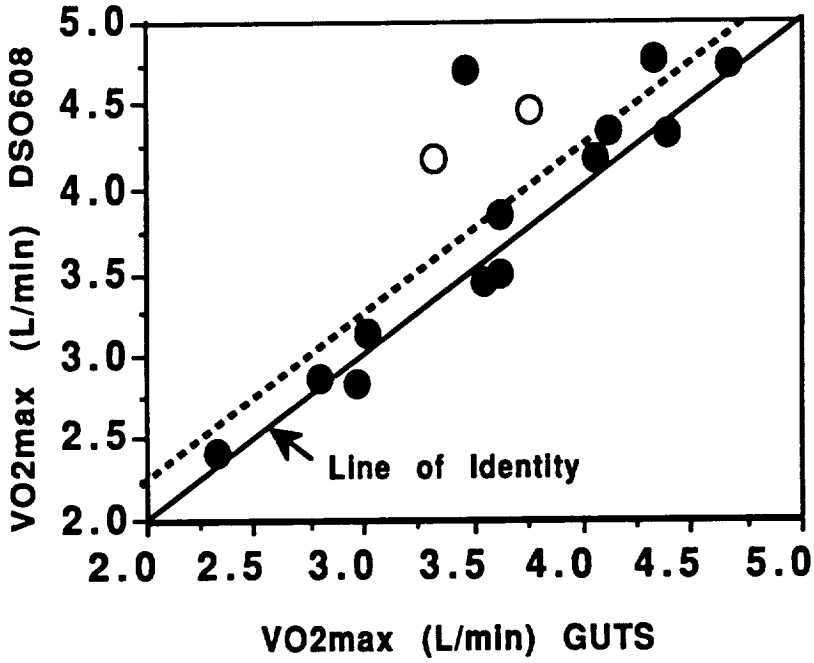


Figure 1: Comparing VO2max between the DSO608 and GUTS protocol. The solid line is the line of identity or a "perfect" match between the protocols. The dotted line is the regression line of GUTS on DSO608. The two open circles represent the two subjects who did not complete the GUTS protocol.

Table 1
Subject Characteristics

Age (yrs)	32.2±5.8
Height (cm)	164.5±27.5
Weight (kg)	73.7±15.1
Gender (M/F)	12/3

REFERENCES

1. Medbø, J. I., Mohn, A. C., Tabata, I., Bahr, R., Vaage, O., & Sejersted, O. M. *Anaerobic Capacity Determined by Maximal Accumulated O2 Deficit.* Journal of Applied Physiology, vol. 64, 1988, pp. 50-60.

2. Medbø, J. I., and Tabata, I. *Relative Importance of Aerobic and Anaerobic Energy Release During Short-lasting Exhausting Bicycle Exercise.* Journal of Applied Physiology, vol. 67(5), 1989, pp. 1881-1886.

3. Rodgers, S., Griffin, C., & Siconolfi, S. *Assessing Anaerobic Capacity on a Treadmill.* (Submitted Abstract to the Aerospace Medicine Association Annual Scientific Meeting for 1993).

4. Schnabel, A. and Kindermann, W: *Assessment of Anaerobic Capacity in Runners.* European Journal of Applied Physiology and Occupational Physiology, vol 52, 1983, pp 42-46.

5. Scott, C.B., Roby, F. B., Lohman, T. G., & Bunt, J.C. *The Maximally Accumulated Oxygen Deficit as an Indicator of Anaerobic Capacity.* Medicine and Science in Sports and Exercise.,vol. 23(5), 1991, pp. 618-624.

6. Siconolfi, S.F., Moore, Jr., A.D. Lane, H. W. *DSO608 The Effects of Microgravity on Aerobic and Anaerobic Energy Expenditure: The Role of Body Composition.* Flight Projects Office, NASA, Johnson Space Center, Houston, TX 77058.

N93-26078

A Hybrid Multigrid Technique for Computing Steady-State Solutions to Supersonic Flows.

Final Report
NASA/ASEE Summer Faculty Fellowship Program – 1992
Johnson Space Center

Prepared by:	Richard Sanders
Academic Rank:	Associate Professor
University & Rank:	University of Houston Department of Mathematics Houston, Texas 77204-3476
Directorate:	Engineering
Branch:	Aerosciences (EG3)
JSC Colleague:	Chien Li
Date Submitted:	September 21, 1992
Contact Number:	NGT-44-001-800

§1. Introduction. Recently, Li and Sanders [2] have introduced a class of finite difference schemes to approximate generally discontinuous solutions to hyperbolic systems of conservation laws. These equations have the form

$$\frac{\partial}{\partial t}\mathbf{q} + \nabla \cdot \mathbf{g}(\mathbf{q}) + \mathbf{s}(\mathbf{q}) = 0$$
$$\mathbf{q}(0) = \mathbf{q}_0,$$

together with relevant boundary conditions. When modelling hypersonic spacecraft reentry, the differential equations above are frequently given by the compressible Euler equations coupled with a nonequilibrium chemistry model. For these applications, steady state solutions are often sought. Many tens (to hundreds) of supercomputer hours can be devoted to a single three space dimensional simulation. The primary difficulty is the inability to rapidly and reliably capture the steady state. In these notes, we demonstrate that a particular variant from the schemes presented in [2] can be combined with a particular multigrid approach to capture steady state solutions to the compressible Euler equations in one space dimension. We show that the rate of convergence to steady state coming from this multigrid implementation is vastly superior to the traditional approach of artificial time relaxation. Moreover, we demonstrate virtual grid independence. That is, the rate of convergence does not depend on the degree of spatial grid refinement.

Before continuing, we review the particular variant of the numerical discretization of $\partial_x \mathbf{g}(\mathbf{q})$ we wish to combine with multigrid. We assume that the problem to be solved is hyperbolic. That is, the Jacobian matrix of $\mathbf{g}(\mathbf{q})$, denoted here by $\partial_{\mathbf{q}}\mathbf{g}(\mathbf{q})$, has real eigenvalues $\lambda_i(\mathbf{q})$, $i = 1, \ldots, m$ and a complete set of eigenvectors $\mathbf{r}_i(\mathbf{q})$. The approximation of $\partial_x \mathbf{g}(\mathbf{q})$ in grid cell j is given by

$$\frac{1}{\Delta x}\left(\mathbf{h}_{\mathbf{g}}(\mathbf{q}_{j+1/2}; \mathbf{q}^l_{j+1/2}, \mathbf{q}^r_{j+1/2}) - \mathbf{h}_{\mathbf{g}}(\mathbf{q}_{j-1/2}; \mathbf{q}^l_{j-1/2}, \mathbf{q}^r_{j-1/2})\right),$$

where $\mathbf{h}_{\mathbf{g}}$ is a two point numerical flux function consistent to $\mathbf{g}$. That is, $\mathbf{h}_{\mathbf{g}}(\mathbf{p}; \mathbf{q}, \mathbf{q}) = \mathbf{g}(\mathbf{q})$. The particular numerical flux function we take below is of Roe type [1]

$$\mathbf{h}_{\mathbf{g}}(\mathbf{p}; \mathbf{a}, \mathbf{b}) = \frac{1}{2}\left(\mathbf{g}(\mathbf{a}) + \mathbf{g}(\mathbf{b}) - R(\mathbf{p})|\Lambda(\mathbf{p})|L(\mathbf{p})(\mathbf{b} - \mathbf{a})\right),$$

where $R(\mathbf{p})$ is the matrix of right eigenvectors to $\partial_{\mathbf{q}}\mathbf{g}(\mathbf{p})$, $L(\mathbf{p})$ the inverse to $R(\mathbf{p})$ and $|\Lambda(\mathbf{p})| = diag(|\lambda_i(\mathbf{p})| \vee \delta)$. We let $\mathbf{q}_{j+1/2} = \frac{1}{2}(\mathbf{q}_{j+1} + \mathbf{q}_j)$ and compute $\mathbf{q}^l_{j+1/2}$ and $\mathbf{q}^r_{j+1/2}$ according to the following recipe.

(i) At each cell interface $j + 1/2$, compute

$$D^{1,0} = L(\mathbf{q}_{j+1/2})(\mathbf{q}_{j+1} - \mathbf{q}_j)$$
$$D^{2,+} = L(\mathbf{q}_{j+1/2})(\mathbf{q}_{j+2} - 2\mathbf{q}_{j+1} + \mathbf{q}_j)$$
$$D^{2,-} = L(\mathbf{q}_{j+1/2})(\mathbf{q}_{j+1} - 2\mathbf{q}_j + \mathbf{q}_{j-1})$$

and componentwise ($i = 1, \ldots, m$)

$$(c_{j+1/2})_i = \begin{cases} \min(|D_i^{2,-}|, 3|D_i^{1,0}|)\text{sign}(D_i^{2,-}) & \text{if } \lambda_i(\mathbf{q}_{j+1/2}) > 0 \\ \min(|D_i^{2,+}|, 3|D_i^{1,0}|)\text{sign}(D_i^{2,+}) & \text{otherwise} \end{cases},$$

to obtain interface values

$$\mathbf{q}^m_{j+1/2} = \mathbf{q}_{j+1/2} - \frac{1}{6}R(\mathbf{q}_{j+1/2})\mathbf{c}_{j+1/2}.$$

(ii) Limit the interface values by

$$\begin{aligned} \mathbf{q}^l_{j+1/2} &= \mathbf{q}_j \\ &\quad + R(\mathbf{q}_j)\left(\text{minmod}(L(\mathbf{q}_j)(\mathbf{q}^m_{j+1/2} - \mathbf{q}_j), \rho L(\mathbf{q}_j)(\mathbf{q}^m_{j-1/2} - \mathbf{q}_j))\right) \\ \mathbf{q}^r_{j+1/2} &= \mathbf{q}_{j+1} \\ &\quad + R(\mathbf{q}_{j+1})\left(\text{minmod}(L(\mathbf{q}_{j+1})(\mathbf{q}^m_{j+1/2} - \mathbf{q}_{j+1}), \rho L(\mathbf{q}_{j+1})(\mathbf{q}^m_{j+3/2} - \mathbf{q}_{j+1}))\right) \end{aligned}$$

where $\rho > 1$ is called a *compression factor.*

Note that for ρ taken larger than 1 we have generically that $\mathbf{q}^l_{j+1/2} = \mathbf{q}^r_{j+1/2} = \mathbf{q}^m_{j+1/2}$. Therefore, generically

$$\mathbf{h}_\mathbf{g}(\mathbf{q}_{j+1/2}; \mathbf{q}^l_{j+1/2}, \mathbf{q}^r_{j+1/2}) = \mathbf{g}(\mathbf{q}^m_{j+1/2}).$$

In the case when $\mathbf{g}(\mathbf{q}) = \mathbf{q}$ and $\mathbf{q} \in \mathbf{R}^1$, the finite difference formula above reduces generically to

$$(\frac{\partial}{\partial x}\mathbf{q})_j \approx \frac{1}{\Delta x}\left(\frac{1}{2}(\mathbf{q}_{j+1} - \mathbf{q}_{j-1}) - \frac{1}{6}(\mathbf{q}_{j+1} - 3\mathbf{q}_j + 3\mathbf{q}_{j-1} - \mathbf{q}_{j-2})\right).$$

(See Section 3.2.)

In the next section we motivate the basics of multigrid and study its convergence for an upwind scheme. In Section 3 a hybrid of multigrid with an approximate Newton's iteration is introduced and analysed. In the last section, we numerically treat an example supersonic flow problem using the difference formula given above together with a multigrid algorithm developed below.

§2.1 The basics of multigrid. Textbooks are filled with iterative techniques to solve large, sparse linear systems. Such systems are generally encountered when seeking approximations to steady-state solutions of partial differential equations in one, two or three space dimensions. The rate of convergence offered by almost all iterative techniques unfortunately depend on the size of the system. Specifically, by increasing the number of spatial grid points, the iterative technique will require a greater number of iterations to achieve a given error reduction. This phenomena can be exhibited by considering the finite difference scheme

$$\frac{1}{\Delta x}(u_j - u_{j-1}) + f(x_j) = 0 \quad j = 1, 2, \ldots, J,$$
$$\frac{1}{\Delta x} = J,$$

with periodic boundary conditions,

$$u_0 = u_J,$$

used to approximate the solution to the differential equation

$$\frac{\partial}{\partial x}u + f(x) = 0,$$
$$u(0) = u(1).$$

(The finite difference scheme above has a nonunique solution provided that $\sum_{j=1}^{J} f(x_j) = 0$.) Letting $\mathbf{u}$ represent the J dimensional vector of unknowns u_j, and $\mathbf{f}$ the J dimensional vector of $f(x_j)$, we can write the finite difference scheme above symbolically as

$$D_J\mathbf{u} + \mathbf{f} = 0, \tag{2.1}$$

where D_J is a $J \times J$ matrix, and we will consider the rate of convergence of the artificial time relaxation technique

$$\mathbf{u}^k = \mathbf{u}^{k-1} - \rho(D_J\mathbf{u}^{k-1} + \mathbf{f}), \tag{2.2}$$
$$\mathbf{u}^0 = \mathbf{v},$$

where $\rho > 0$ is a relaxation parameter and $\mathbf{v}$ is an arbitrary starting vector. Clearly, if $\lim_{k\to\infty} \mathbf{u}^k$ exists, the limit solves (2.1).

Recall the discrete Fourier transform of a J-periodic grid function v_j:

$$\hat{v}_l = \frac{1}{\sqrt{J}} \sum_{j=1}^{J} v_j e^{\frac{2\pi i}{J} lj},$$

and its associated inversion formula,

$$v_j = \frac{1}{\sqrt{J}} \sum_{l=1}^{J} \hat{v}_l e^{-\frac{2\pi i}{J} lj}.$$

Let $\mathbf{u}^k$ denote the *kth* iterate coming from (2.2), let $\mathbf{u}$ denote the unique stationary solution to (2.1) normalized so that $\sum_{j=1}^{J} u_j = \sum_{j=1}^{J} u_j^0$, and let $\epsilon^k = \mathbf{u} - \mathbf{u}^k$ denote the artificial time iteration error after k iterations. By taking

$$\epsilon_j^k = \frac{1}{\sqrt{J}} \sum_{l=1}^{J} \hat{\epsilon}_l^k e^{-\frac{2\pi i}{J} lj},$$

one easily calculates that the the *lth* Fourier coefficient of error at iteration step k is given explicitly by

$$\hat{\epsilon}_l^k = \left((1 - \frac{\rho}{\Delta x}) + \frac{\rho}{\Delta x} e^{\frac{2\pi i}{J} l} \right)^k \hat{\epsilon}_l^0.$$

(Note that the normalization of $\mathbf{u}$ implies $\hat{\epsilon}_J^0 = 0$.) Taking $\sigma \equiv \rho/\Delta x < 1$, we have that

$$|(1 - \frac{\rho}{\Delta x}) + \frac{\rho}{\Delta x} e^{\frac{2\pi i}{J} l}| < 1,$$

for each $l = 1, 2, \ldots, J-1$. Moreover, the high frequency components of the error are contained in the Fourier coefficients $\hat{\epsilon}_l^k$ with $l \approx J/2$. From above, we see that these high frequency components decay at a rate on the order of

$$|1 - 2\sigma|.$$

Therefore, these coefficients decay at a rate that is independent of the number of grid points J. On the other hand, the low frequency components of the error are contained in $\hat{\epsilon}_l^k$ with $l \approx 1$ or with $l \approx J-1$. These components decay at a rate on the order of

$$1 - \frac{2\pi^2}{J^2} \sigma(1 - \sigma).$$

Therefore, for this example, doubling the number of grid points J can necessitate many more than twice the number of artificial time iterations to achieve a given error reduction.

The basic idea of multigrid is the following: One attempts to capture the low frequency components of $\mathbf{u}$ on a coarse mesh. This can be accomplished by a variety of methods. The coarse grid information is then included into the current guess on the fine grid, and the high frequency components of the error can be rapidly damped there by using an iterative technique such as the one above. Normally, the divide and conquer strategy is employed in multigrid. That is, a nested sequence of coarse meshes is utilized to rapidly capture all frequencies of the solution.

To illustrate the multigrid approach, we will apply an *ideal* two grid multigrid strategy to example (2.1) above. Given the current approximation $\mathbf{u}^k$ to $\mathbf{u}$, we have that

$$D_J\mathbf{u} + \mathbf{f} = D_J(\mathbf{u}^k + \epsilon^k) + \mathbf{f} = 0$$

or,

$$D_J\epsilon^k + \mathbf{r}^k = 0, \tag{2.3}$$

where the residual is defined by $\mathbf{r}^k = D_J\mathbf{u}^k + \mathbf{f}$, and D_J is the $J \times J$ difference matrix defined above. The solution to the residual equation (2.3) is approximated by solving

$$D_{J/2}\epsilon^k_{J/2} + \mathbf{r}^k_{J/2} = 0$$

where,

$$\mathbf{r}^k_{J/2} = I_{(J\to J/2)}\mathbf{r}^k.$$

$I_{(J\to J/2)}$ denotes an injection operator that takes vectors from $\mathbf{R}^J$ into the lower dimensional space $\mathbf{R}^{J/2}$. (Solving the coarse grid equation above exactly is what defines an ideal two grid strategy. In practice, multigrid is nested. That is, a nested sequence of lower dimensional multigrid iterations is applied to a nested sequence of coarser grid residual equations.) The fine grid error is then approximated by

$$\epsilon^k \approx I_{(J/2\to J)}\epsilon^k_{J/2} \equiv \epsilon^k_{cg},$$

where $I_{(J/2\to J)}$ denotes an interpolation operator that takes vectors from $\mathbf{R}^{J/2}$ back into $\mathbf{R}^J$. ϵ^k_{cg} is called the *coarse grid correction*. One expects that ϵ^k_{cg} will agree well with the true error in the low frequencies. However, it in general will not agree well with the true error in the high frequency regime. To capture the high frequencies, we may for example use the artificial time scheme above. e.g., for $n = 1, 2, \ldots, \nu$

$$\mathbf{e}^n = \mathbf{e}^{n-1} - \rho(D_J\mathbf{e}^{n-1} + \mathbf{r}^k), \tag{2.4}$$

with

$$\mathbf{e}^0 = \epsilon^k_{cg}.$$

(This step of multigrid is often referred to as the *smoother*. The term smoother is in fact a misnomer. In actuality, this step is used to capture the high frequency components of the error.) Finally, since

$$\mathbf{u} = \mathbf{u}^k + \epsilon^k \approx \mathbf{u}^k + \mathbf{e}^\nu,$$

we update $\mathbf{u}^k$ via

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{e}^\nu.$$

This defines one cycle of a simple two level multigrid technique. We should remark that the update step above is equivalent to setting

$$\mathbf{u}^{k+1} = S^\nu_J(\mathbf{u}^k + \epsilon^k_{cg}, \mathbf{f}),$$

where $S^\nu_J(\mathbf{v}, \mathbf{f})$ denotes ν iterations of the basic iteration scheme (2.2) on grid J.

§2.2 Ideal two grid analysis for the example first order one sided scheme. A multigrid strategy requires three basic ingredients. They are:

(i) The so-called smoother; see (2.4) above.

(ii) The fine to coarse grid residual injection operator; $I_{(J\to J/2)}$ above.

(iii) The coarse to fine grid error interpolation operator; $I_{(J/2\to J)}$ above.

The performance of a multigrid algorithm relies heavily on the choice of these ingredients. In this subsection, we analyse the convergence rate of the strategy outlined above taking the artificial time relaxation technique (2.2) as the smoother, and

$$(I_{(J\to J/2)}\mathrm{r})_j = \frac{1}{2}(r_{2j} + r_{2j-1}), \tag{2.5}$$

as the fine to coarse grid injection operator, and

$$\begin{aligned}(I_{(J/2\to J)}\epsilon)_{2j} &= \epsilon_j\\ (I_{(J/2\to J)}\epsilon)_{2j-1} &= \epsilon_j,\end{aligned} \tag{2.6}$$

as the coarse to fine grid interpolation operator. In the context of finite volume, these inter-grid operators seem most natural. A multigrid iteration step above is written symbolically as

$$\begin{aligned}\epsilon^k_{cg} &= I_{(J/2\to J)}S^\infty_{J/2}(0, I_{(J\to J/2)}\mathrm{r}^k)\\ \mathbf{u}^{k+1} &= S^\nu_J(\mathbf{u}^k + \epsilon^k_{cg}, \mathbf{f}).\end{aligned} \tag{2.7}$$

Using the fact that $\mathbf{u} = S^\nu_J(\mathbf{u}, \mathbf{f})$, we may write (2.7) in terms of the error ϵ^k as

$$\begin{aligned}\epsilon^k_{cg} &= I_{(J/2\to J)}S^\infty_{J/2}(0, -I_{(J\to J/2)}D_J\epsilon^k)\\ \epsilon^{k+1} &= S^\nu_J(\epsilon^k - \epsilon^k_{cg}, 0),\end{aligned} \tag{2.8}$$

where again we normalize $\mathbf{u}$ such that $\sum_{j=1}^J (u_j - u^0_j) \equiv \sum_{j=1}^J \epsilon^0_j = 0$.

Unfortunately, unlike the single grid example above, the two grid multigrid algorithm (2.8) does not completely decouple into Fourier modes. Following [3], a J-periodic grid function $\mathbf{v}$ can be decomposed into a Fourier series on the odd grid points, and a Fourier series on the even grid points. The explicit formulae are

$$v_j = \begin{cases} \dfrac{1}{\sqrt{J/2}}\displaystyle\sum_{l=1}^{J/2} \tilde{v}_l e^{-\frac{2\pi i}{J/2}l(j+1)/2} & \text{if } j = 1, 3, \ldots, J-1\\[2ex] \dfrac{1}{\sqrt{J/2}}\displaystyle\sum_{l=1}^{J/2} \tilde{\tilde{v}}_l e^{-\frac{2\pi i}{J/2}lj/2} & \text{if } j = 2, 4, \ldots, J\end{cases} \tag{2.9}$$

In fact, if $v_j = \frac{1}{\sqrt{J}} \sum_{l=1}^{J} \hat{v}_l e^{-\frac{2\pi i}{J} lj}$, a simple calculation will reveal that

$$\tilde{v}_l = \frac{\sqrt{2}}{2} \left(\hat{v}_l - \hat{v}_{l+J/2} \right) e^{\frac{2\pi i}{J} l}, \tag{2.10}$$

$$\tilde{\tilde{v}}_l = \frac{\sqrt{2}}{2} \left(\hat{v}_l + \hat{v}_{l+J/2} \right),$$

for each $l = 1, 2, \ldots, J/2$.

(2.9) and (2.10) allow us to examine the degree to which the coarse grid correction ϵ_{cg}^k approximates the true error ϵ^k. Writing

$$\epsilon_j^k = \frac{1}{\sqrt{J}} \sum_{l=1}^{J} \hat{\epsilon}_l^k e^{-\frac{2\pi i}{J} lj},$$

or equivalently

$$\epsilon_j^k = \begin{cases} \frac{1}{\sqrt{J/2}} \sum_{l=1}^{J/2} \tilde{\epsilon}_l^k e^{-\frac{2\pi i}{J/2} l(j+1)/2} & \text{if } j = 1, 3, \ldots, J-1 \\ \frac{1}{\sqrt{J/2}} \sum_{l=1}^{J/2} \tilde{\tilde{\epsilon}}_l^k e^{-\frac{2\pi i}{J/2} lj/2} & \text{if } j = 2, 4, \ldots, J \end{cases},$$

we compute that

$$\widetilde{D_J \epsilon^k}_l = \frac{1}{\Delta x} \left(\tilde{\epsilon}_l^k - \tilde{\tilde{\epsilon}}_l^k e^{\frac{2\pi i}{J/2} l} \right),$$

$$\widetilde{\widetilde{D_J \epsilon^k}}_l = \frac{1}{\Delta x} \left(\tilde{\tilde{\epsilon}}_l^k - \tilde{\epsilon}_l^k \right).$$

Therefore, on the coarse grid, with j running from 1 to $J/2$, we have

$$\begin{aligned} (I_{(J \to J/2)} D_J \epsilon^k)_j &= \frac{1}{2} \left((D_J \epsilon^k)_{2j-1} + (D_J \epsilon^k)_{2j} \right) \\ &= \frac{1}{\sqrt{J/2}} \sum_{l=1}^{J/2} \frac{1}{2} \left(\widetilde{D_J \epsilon^k}_l + \widetilde{\widetilde{D_J \epsilon^k}}_l \right) e^{-\frac{2\pi i}{J/2} lj} = \frac{1}{\sqrt{J/2}} \sum_{l=1}^{J/2} \frac{1}{2\Delta x} (1 - e^{\frac{2\pi i}{J/2} l}) \tilde{\tilde{\epsilon}}_l^k e^{-\frac{2\pi i}{J/2} lj}. \end{aligned}$$

Solving the coarse grid problem exactly leads to

$$S_{J/2}^{\infty}(0, -I_{(J \to J/2)} D_J \epsilon^k) = \frac{1}{\sqrt{J/2}} \sum_{l=1}^{J/2} \tilde{\tilde{\epsilon}}_l^k e^{-\frac{2\pi i}{J/2} lj},$$

and we find, after applying the coarse to fine grid interpolation (2.6), that the coarse grid correction is given by

$$(\epsilon_{cg}^k)_{2j-1} = (\epsilon_{cg}^k)_{2j} = \frac{1}{\sqrt{J/2}} \sum_{l=1}^{J/2} \tilde{\tilde{\epsilon}}_l^k e^{-\frac{2\pi i}{J/2} lj}.$$

Making use of the formulae in (2.10), we can write the coarse grid correction in terms of $\{\hat{\epsilon}_l^k\}_{l=1}^J$

$$(\hat{\epsilon}_{cg}^k)_l = \frac{1}{2}(1 + e^{-\frac{2\pi i}{J}l})(\hat{\epsilon}_l^k + \hat{\epsilon}_{l+J/2}^k) \tag{2.11a}$$

$$(\hat{\epsilon}_{cg}^k)_{l+J/2} = \frac{1}{2}(1 - e^{-\frac{2\pi i}{J}l})(\hat{\epsilon}_{l+J/2}^k + \hat{\epsilon}_l^k) \tag{2.11b}$$

$$(l = 1, 2, \ldots, J/2).$$

Notice that for $l \approx 1$, the coarse grid correction coefficients given in (2.11a) agree well with the true error coefficients modulo a high frequency component. The same conclusion follows from (2.11b) with $l + J/2 \approx J$. That is, modulo high frequency pollution, the coarse grid correction accurately predicts the low frequency components of the error.

Finally, after applying ν iterations of the smoother ($\epsilon^{k+1} = S_J^\nu(\epsilon^k - \epsilon_{cg}^k, 0)$), one cycle of this ideal multigrid iteration yields an error which satisfies the coupled system

$$\begin{pmatrix} \hat{\epsilon}_l^{k+1} \\ \hat{\epsilon}_{l+J/2}^{k+1} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} a_l^\nu(1 - e^{-\frac{2\pi i}{J}l}) & -a_l^\nu(1 + e^{-\frac{2\pi i}{J}l}) \\ -a_{l+J/2}^\nu(1 - e^{-\frac{2\pi i}{J}l}) & a_{l+J/2}^\nu(1 + e^{-\frac{2\pi i}{J}l}) \end{pmatrix} \begin{pmatrix} \hat{\epsilon}_l^k \\ \hat{\epsilon}_{l+J/2}^k \end{pmatrix}, \tag{2.12}$$

where

$$a_l^\nu = ((1 - \sigma) + \sigma e^{\frac{2\pi i}{J}l})^\nu.$$

We calculate the eigenvalues of the amplification matrix above:

$$\lambda_1 = 0$$
$$\lambda_2 = \frac{1}{2}(a_l^\nu(1 - e^{-\frac{2\pi i}{J}l}) + a_{l+J/2}^\nu(1 + e^{-\frac{2\pi i}{J}l})).$$

In the special case when $\nu = 1$ and $\sigma = 1/2$, we have that $\lambda_2 = 0$. Therefore, this multigrid scheme yields the exact solution u after at most two cycles regardless of its initial guess. More generally for any $\nu \geq 1$ and any fixed $0 < \sigma < 1$, one can show, independently of the number of grid points J, that $|\lambda_2|$ is strictly less than one. So in the general case, the rate of convergence of this ideal multigrid scheme does not degrade with grid size.

As mentioned above, the ideal scheme is not practical. One does not wish to solve the coarse grid problem exactly. Nevertheless, the analysis above does indicate that the fully nested multigrid approach should converge rapidly in a grid size independent manner.

§3.1 A hybrid of multigrid. Consider for the moment the possibly nonlinear differential equation

$$\frac{\partial}{\partial x} g(u) + f(x) = 0,$$

together with given boundary conditions. Assuming that this equation has a solution, Newton's method is developed by setting $u = u^k + \epsilon^k$, where u^k is the current guess to the solution, and expanding $g(u^k + \epsilon^k) = g(u^k) + g'(u^k)\epsilon^k + \cdots$. Therefore,

$$0 = \frac{\partial}{\partial x} g(u) + f(x) \approx \frac{\partial}{\partial x}(g(u^k) + g'(u^k)\epsilon^k) + f(x),$$

and we let ϵ^k_{nm} solve the linearized equation

$$\frac{\partial}{\partial x}(g'(u^k)\epsilon^k_{nm}) + \frac{\partial}{\partial x} g(u^k) + f(x) = 0,$$

to update

$$u^{k+1} = u^k + \epsilon^k_{nm}.$$

This idea extends equally well to finite difference schemes associated to this differential equation. That is, if we seek the solution to the finite difference scheme

$$D_J(g; \mathbf{u}) + \mathbf{f} = 0,$$

where $D_J(g; \mathbf{u})$ denotes a finite difference operator consistent to $\partial_x g(u)$, we solve

$$\partial_{\mathbf{u}}(D_J(g; \mathbf{u}^k))\epsilon^k_{nm} + D_J(g; \mathbf{u}^k) + \mathbf{f} = 0,$$

where $\partial_{\mathbf{u}}(D_J(g; \mathbf{u}^k))$ denotes the Jacobian matrix of $D_J(g; \mathbf{u})$ at $\mathbf{u} = \mathbf{u}^k$, and update

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \epsilon^k_{nm}.$$

Newton's method has two main drawbacks: (i) Each iteration requires the inversion of a large linear system. (ii) The Jacobian matrix of $D_J(g; \mathbf{u}))$ may be quite complicated to compute analytically. In fact, for many modern finite difference schemes, the function $D_J(g; \mathbf{u}))$ may lack the necessary smoothness to gain the quadratic convergence offered by Newton's method. We propose the following simplification: Suppose we wish to solve

$$D^{(hi)}_J(g; \mathbf{u}) + \mathbf{f} = 0,$$

where $D^{(hi)}_J(g; \mathbf{u})$ denotes a (possibly high order) finite difference operator consistent with $\partial_x(g(u))$. Let $D^{(lo)}_J(g'(u))\epsilon$ denote a (possibly low order) finite difference operator that is consistent to $\partial_x(g'(u)\epsilon)$. Rather than solving the correct linearization of the finite difference scheme above, we solve (or approximate the solution of)

$$D^{(lo)}_J(g'(u^k))\tilde{\epsilon}^k_{nm} + D^{(hi)}_J(g; \mathbf{u}^k) + \mathbf{f} = 0, \tag{3.1}$$

to update $\mathbf{u}^{k+1} = \mathbf{u}^k + \epsilon^k_{\underset{nm}{\sim}}$. Certainly there is no reason to expect quadratic convergence from this approach. Therefore, rather than solving (3.1) exactly, we approximate $\epsilon^k_{\underset{nm}{\sim}}$ by applying multigrid to (3.1). Specifically, suppose that $\mathbf{u}^k$ is known. Then

(3.2)

Step 1: Compute $R(\mathbf{u}^k) = D_J^{(hi)}(g;\mathbf{u}^k) + \mathbf{f}$.

Step 2: Apply $\eta \geq 1$ cycles of multigrid with initial guess 0 to approximate the solution to

$$D_J^{(lo)}(g'(u^k))\epsilon^k_{\underset{nm}{\sim}} + R(\mathbf{u}^k) = 0,$$

calling the result ϵ^k_{mg}.

Step 3: Update $\mathbf{u}^{k+1} = \mathbf{u}^k + \epsilon^k_{mg}$.

Step 4: (Optional) Steps 1-3 may not capture the high frequency components of $\mathbf{u}$ well. Therefore, applying a few iteration of a smoother consistent to $D_J^{(hi)}(g;\mathbf{u}) + \mathbf{f} = 0$ may be advised; (see below).

§3.2 Convergence of the hybrid multigrid approach for a model problem. Whether iterates coming from (3.2) converge or not depends in a crucial way on the choice of both $D_J^{(lo)}$ and $D_J^{(hi)}$. We again only perform ideal two grid multigrid analysis to the linear problem $g(u) = u$. The problem is made somewhat more interesting by taking an upwind third order scheme for $D_J^{(hi)}$

$$(3.3a) \qquad (D_J^{(hi)}\mathbf{u})_j = \frac{1}{2\Delta x}(u_{j+1} - u_{j-1}) - \frac{1}{6\Delta x}(u_{j+1} - 3u_j + 3u_{j-1} - u_{j-2})$$

and a first order upwind scheme for $D_J^{(lo)}$

$$(3.3b) \qquad (D_J^{(lo)}\epsilon)_j = \frac{1}{\Delta x}(\epsilon_j - \epsilon_{j-1}),$$

together with periodic boundary conditions. (Note that (3.3a) is third order in the sense of cell averages.) We still require that $\sum_{j=1}^J f(x_j) = 0$ for solvability.

Before studying (3.1) & (3.3) hybridized with multigrid, observe that if $\epsilon^k = \mathbf{u} - \mathbf{u}^k$ (where again $\mathbf{u}$ is normalized such that $\sum_{j=1}^J (u_j - u_j^0) = 0$), then $\epsilon^k_{\underset{nm}{\sim}}$ (see 3.1) satisfies

$$(3.4) \qquad (\hat{\epsilon^k_{\underset{nm}{\sim}}})_l = \psi_l \hat{\epsilon}^k_l,$$

where the amplification factor ψ_l is given by

$$\psi_l = \frac{(-i\sin(\theta_l) - 1/3(\cos(\theta_l) - 1)(1 - e^{i\theta_l}))}{(1 - e^{i\theta_l})}, \qquad \theta_l = \frac{2\pi}{J}l.$$

The error reduction rate expected from algorithm (3.1) (or equivalently algorithm (3.2) with Step 2 solving exactly) is on the order of $\max_{1\le l\le J-1}|1-\psi_l| \approx 0.53$ and the maximum is attained in the mid frequencies. For the high frequencies $l \approx J/2$ we have that $|1-\psi_l| \approx 1/3$, and for the low frequencies $l \approx 1$ or $l \approx J-1$ we have that $|1-\psi_l| \approx 0$. So we see that the low frequency components of the error are captured quite rapidly whereas the mid to high frequencies decay less rapidly. Step 4 can therefore be utilized to knock down the mid to high frequency components of the error without affecting the low frequency convergence.

The results of Section 2 make the analysis of the hybrid multigrid algorithm (3.2) applied to the difference operators in (3.3) an easy matter. Since $\epsilon^k_{\underset{nm}{\sim}} - \epsilon^k_{mg}$ is the error of Step 2, using (2.12) we find that

$$\begin{pmatrix} (\widehat{\epsilon}^k_{\underset{nm}{\sim}} - \widehat{\epsilon}^k_{mg})_l \\ (\widehat{\epsilon}^k_{\underset{nm}{\sim}} - \widehat{\epsilon}^k_{mg})_{l+J/2} \end{pmatrix}$$
$$= \left[\frac{1}{2}\begin{pmatrix} a^\nu_l(1-e^{-\frac{2\pi i}{J}l}) & -a^\nu_l(1+e^{-\frac{2\pi i}{J}l}) \\ -a^\nu_{l+J/2}(1-e^{-\frac{2\pi i}{J}l}) & a^\nu_{l+J/2}(1+e^{-\frac{2\pi i}{J}l}) \end{pmatrix}\right]^\eta \begin{pmatrix} (\widehat{\epsilon}^k_{\underset{nm}{\sim}})_l \\ (\widehat{\epsilon}^k_{\underset{nm}{\sim}})_{l+J/2} \end{pmatrix}.$$

So if we denote the 2×2 multigrid amplification matrix above by $\widehat{M}_l^{(\sigma,\nu,\eta)}$, and define

$$\widehat{\Psi}_l = \begin{pmatrix} \psi_l & 0 \\ 0 & \psi_{l+J/2} \end{pmatrix},$$

then referring back to (3.4) we find that the 2×2 matrices

$$\widehat{A}_l = \left(I - (I - \widehat{M}_l^{(\sigma,\nu,\eta)})\widehat{\Psi}_l\right), \qquad (l = 1,2,\ldots,J/2), \tag{3.5}$$

define the amplification matrix to the full (Steps 1-3) hybrid algorithm (3.2). The spectral radius of the hybrid amplification factor is plotted as a function of σ in Figure 1. (Recall that σ is the ratio of the artificial time step size ρ to the grid size Δx in the approximate linearization multigrid smoother.) Note that for $\sigma = 0.40$ we actually achieve better error reduction using multigrid in Step 2 than if we solved (3.1) exactly.

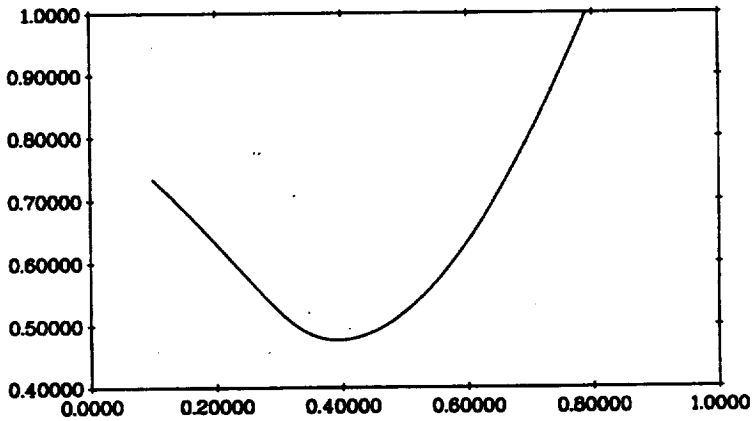


Figure 1. The spectral radius of $diag(\hat{A}_l)$ plotted against σ. ($\nu = 2, \eta = 2.$)

Notice the sensitivity of the error reduction rate to σ in Figure 1. This clearly indicates the relationship of multigrid error reduction to linearization wave speeds. Therefore, for nonlinear problems where the wave speeds are not constant, (especially for nonlinear systems with multiple wave speeds), we expect a poor convergence rate from the primitive algorithm outlined above. This problem will be addressed in the next section.

§4. Application to the one dimensional Euler equations. The partial differential equations that govern the flow of a compressible and inviscid gas in a quasi one dimensional expanding duct are

$$\begin{aligned}
&\frac{\partial}{\partial t}(\rho A) + \frac{\partial}{\partial x}(\rho u A) = 0 \\
&\frac{\partial}{\partial t}(\rho u A) + \frac{\partial}{\partial x}\left((\rho u^2 + p)A\right) - p\frac{\partial}{\partial x}A = 0 \\
&\frac{\partial}{\partial t}(\rho e A) + \frac{\partial}{\partial x}\left((\rho e + p)uA\right) = 0,
\end{aligned}$$

where ρ is the fluid's density, u its velocity and e its total energy per unit mass. The given function $A = A(x)$ defines the cross sectional area of the duct as a function of position along its length. p represents the fluid's pressure and is given by the equation of state $p = (\gamma - 1)\rho(e - u^2/2)$. We take $\gamma = 1.4$ in the results presented below. We seek a steady state solution to problem on the interval $0 \le x \le 10$, taking a supersonic inflow boundary condition at $x = 0$: $(\rho, u, p) = (0.502, 1.299, 0.3809)$; and a subsonic outflow condition at $x = 10$: $(\rho, u, p) = (0.776, 0.5063, 0.7475)$. The cross sectional area is given by $A(x) = 1.398 + 0.347\tanh(0.8x - 4)$.

To demonstrate a need for a technique that speeds convergence to steady state, we seek a steady state to this duct example by applying the third order accurate finite difference scheme from the introduction together with artificial time relaxation. Specifically, we iterate $k = 1, 2, \ldots$

$$\mathbf{q}^k = \mathbf{q}^{k-1} - \rho R^{(hi)}(\mathbf{q}^{k-1}),$$

where the residual $R^{(hi)}(\mathbf{q})$ is given by

$$(R^{(hi)}(\mathbf{q}))_j = \frac{1}{\Delta x}\left(\mathbf{h_g}(\mathbf{q}_{j+1/2}; \mathbf{q}^l_{j+1/2}, \mathbf{q}^r_{j+1/2}) - \mathbf{h_g}(\mathbf{q}_{j-1/2}; \mathbf{q}^l_{j-1/2}, \mathbf{q}^r_{j-1/2})\right) + \mathbf{s}(\mathbf{q}_j).$$

The compression factor in the slope limiting above as well as in all other examples presented is taken to be 3. We initialize the iteration by a linear interpolation of the boundary data in the conserved variables. The relaxation factor ρ is taken so that the CFL number (the ratio of time step size times the fastest wave speed to the space step size) is 0.30 and we

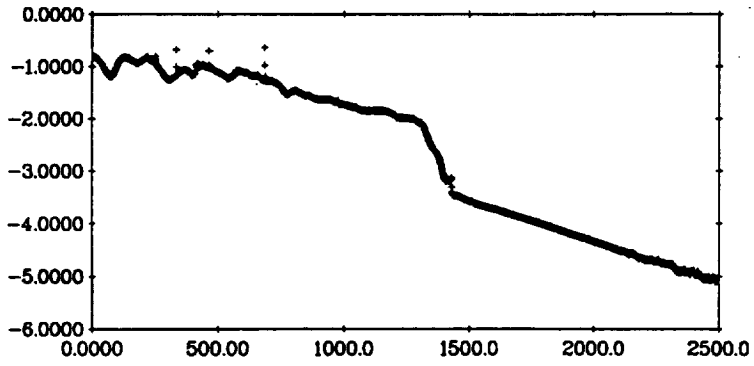


Figure 2. $\frac{1}{2}\log(\frac{1}{64}\sum_{j=1}^{64}(R_1^{(hi)}(q^k))_j^2)$ plotted against iterations for the artificial time scheme.

use 64 grid points on the interval $[0,10]$. In Figure 2, we plot the base 10 log of the L_2 norm of the (first component of the) residual versus the number of iterations.

To apply the multigrid algorithm (3.2), we need to decide on the approximate linearization of the third order scheme used above. We take for $D^{(lo)}(g'(\mathbf{q}))\epsilon$ the following first order scheme

$$\frac{1}{\Delta x}(\mathbf{h}^{(lo)}_{\partial_{\mathbf{q}}\mathbf{g}}(\mathbf{q}_{j+1/2};\epsilon_j,\epsilon_{j+1}) - \mathbf{h}^{(lo)}_{\partial_{\mathbf{q}}\mathbf{g}}(\mathbf{q}_{j-1/2};\epsilon_{j-1},\epsilon_j)),$$

where

$$\begin{aligned}&\mathbf{h}^{(lo)}_{\partial_{\mathbf{q}}\mathbf{g}}(\mathbf{q}_{j+1/2};\epsilon_j,\epsilon_{j+1})\\ &\quad=\frac{1}{2}(\partial_{\mathbf{q}}\mathbf{g}(\mathbf{q}_{j+1/2})(\epsilon_{j+1}+\epsilon_j) - |\partial_{\mathbf{q}}\mathbf{g}(\mathbf{q}_{j+1/2})|(\epsilon_{j+1}-\epsilon_j)).\end{aligned}$$

As usual, $|\partial_{\mathbf{q}}\mathbf{g}(\mathbf{q})|$ denotes the matrix $R(\mathbf{q})|\Lambda(\mathbf{q})|L(\mathbf{q})$. The obvious choice for a smoothing iteration scheme is

$$\begin{aligned}\epsilon_j^{k+1} = \epsilon_j^k - \rho\Big(&\frac{1}{\Delta x}(\mathbf{h}^{(lo)}_{\partial_{\mathbf{q}}\mathbf{g}}(\mathbf{q}_{j+1/2};\epsilon_j^k,\epsilon_{j+1}^k)\\ &- \mathbf{h}^{(lo)}_{\partial_{\mathbf{q}}\mathbf{g}}(\mathbf{q}_{j-1/2};\epsilon_{j-1}^k,\epsilon_j^k)) + \partial_{\mathbf{q}}\mathbf{s}(\mathbf{q}_j)\epsilon_j^k + (R^{(hi)}(\mathbf{q}))_j\Big),\end{aligned}$$

where ρ is fixed according to the largest wave speed of the variable coefficient problem. However, from our analysis in Section 3, we can expect this approach to yield very poor convergence rates. We base this conclusion on the following observation: The choice of ρ for the smoother above must be based on the largest eigenvalue of $\partial_{\mathbf{q}}\mathbf{g}$. This is needed for stability. Since the flow contains a supersonic region, the eigenvalue $u+c$ (c is the local speed of sound and $u > c$ is assumed) is quite large. Therefore, ρ must be taken quite small. On the other hand, near a sonic region, $u-c$ is quite small. Therefore with this simple smoother, the hybrid multigrid performance in sonic regions can be predicted by the results in Figure 1 with σ near zero. There, the error reduction rate is near unity.

For this reason, we propose characteristic time stepping. This is accomplished by a simple modification of the smoothing iteration above:

$$\epsilon_j^{k+1} = \epsilon_j^k - \rho R(\mathbf{q}_j)|1/\Lambda(\mathbf{q}_j)|L(\mathbf{q}_j)\Big(\frac{1}{\Delta x}(\mathbf{h}^{(lo)}_{\partial_{\mathbf{q}}\mathbf{g}}(\mathbf{q}_{j+1/2};\epsilon_j^k,\epsilon_{j+1}^k)$$
$$-\mathbf{h}^{(lo)}_{\partial_{\mathbf{q}}\mathbf{g}}(\mathbf{q}_{j-1/2};\epsilon_{j-1}^k,\epsilon_j^k)) + \partial_{\mathbf{q}}\mathbf{s}(\mathbf{q}_j)\epsilon_j^k + (R^{(hi)}(\mathbf{q}))_j\Big),$$

where now, ρ is taken to be $\rho = 0.4\Delta x$.

We implement algorithm (3.2) to this duct test problem, again using 64 grid points on the interval $[0, 10]$. Multigrid Step 2 is done using every grid level 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, with $\nu = 4$ and $\eta = 1$. Step 3 is relaxed somewhat, specifically $\mathbf{q}^{k+1} = \mathbf{q}^k + \frac{1}{2}\epsilon^k_{mg}$, because of the presence of the supersonic-subsonic shock. Around the shock, the slope limiting algorithm introduces a small amount of numerical viscosity and this viscosity can introduce an instability into the approximate Newton iteration; (see equation (3.4)). Step 4 of the hybrid algorithm is also utilized with 2 characteristic time stepping iterations applied to the third order residual.

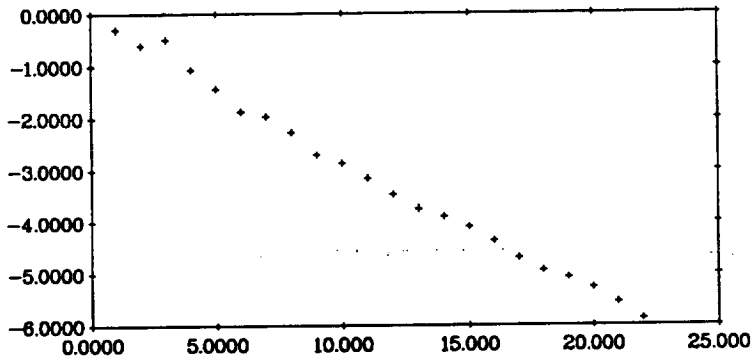


Figure 3. Log residual versus hybrid multigrid cycles, again applied to the duct problem with 64 grid points.

Comparing Figure 2 with Figure 3, we see that the multigrid algorithm reduces the residual of this test problem in 100 times fewer iterations than the artificial time scheme, whereas each multigrid cycle costs on the order of only 4 times the work. Figure 5 depicts the residual reduction from the hybrid multigrid scheme on 5 grids; from level=4 (16 grid points) to level=8 (256 grid points). This last figure demonstrates the virtual grid independence of the hybrid algorithm.

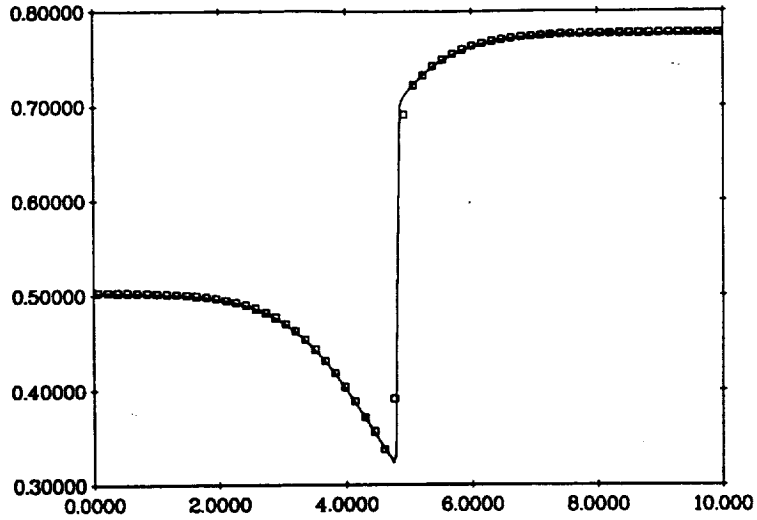


Figure 4. The computed density for the duct test problem using 64 grid points.

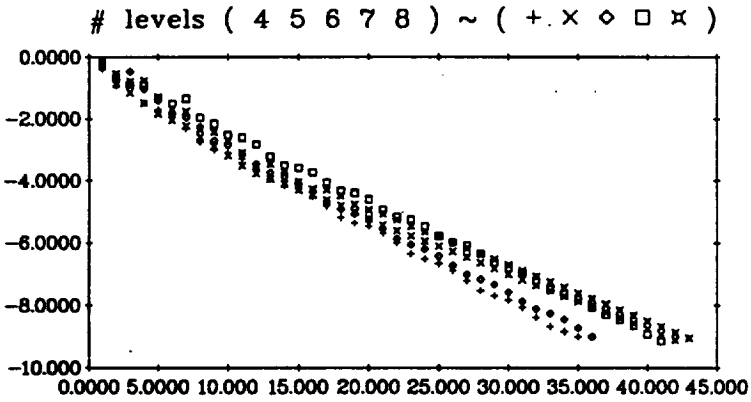


Figure 5. The log residual versus hybrid multigrid cycles for the duct test problem using 16 to 256 grid points.

REFERENCES

[1] P.L. Roe, "Approximate Riemann solvers, parameter vectors, and difference schemes", *J. Comp. Phys.*, 43 (1981), pp. 357-372.

[2] R. Sanders and C.P. Li, "A variation nonexpansive central differencing scheme for nonlinear hyperbolic conservation laws", in Proc. 10th Inter. Conf. Computing Methods in Applied Sciences and Engineering, R. Glowinski ed., Nova Sciences, New York, 1992, pp. 511- 524.

[3] B. Stoufflet, preprint.

N93-26079

MODELING OF THE WSTF FRICTIONAL HEATING APPARATUS IN HIGH PRESSURE SYSTEMS

Final Report

NASA/ASEE Summer Faculty Fellowship Program 1992

Johnson Space Center

Prepared by:	Christopher T. Skowlund, Ph.D.
Academic Rank:	Assistant Professor
University & Department:	New Mexico State University Chemical Engineering Department Las Cruces, New Mexico 88003
NASA/JSC	
Directorate:	Johnson Space Center
Division:	White Sands Test Facility
Branch:	Laboratories Office
JSC Colleague:	Harold D. Beeson, Ph.D.
Date Submitted:	August 11, 1992
Contract Number:	NGT-44-005-803

ABSTRACT

In order to develop a computer program able to model the frictional heating of metals in high pressure oxygen or nitrogen a number of additions have been made to the frictional heating model originally developed for tests in low pressure helium. These additions include: 1) a physical property package for the gases to account for departures from the ideal gas state, 2) two methods for spatial discretization (finite differences with quadratic interpolation or orthogonal collocation on finite elements) which substantially reduce the computer time required to solve the transient heat balance, 3) more efficient programs for the integration of the ordinary differential equations resulting from the discretization of the partial differential equations and 4) two methods for determining the best-fit parameters via minimization of the mean square error (either a direct search multivariable simplex method or a modified Levenburg-Marquardt algorithm). The resulting computer program has been shown to be accurate, efficient and robust for determining the heat flux or friction coefficient vs. time at the interface of the stationary and rotating samples.

INTRODUCTION

The frictional heating apparatus at WSTF is used to determine the flammability of various metals undergoing frictional heating in oxygen-rich environments. A detailed description of this apparatus has been given elsewhere[1] as well as a list of the previous mathematical models for the frictional heating.[2]

As a part of the 1991 NASA/ASEE Summer Faculty Fellowship Program a computer model[2] of the frictional heating apparatus was developed for frictional heating experiments in low pressure helium (approximately 75 psia). The model that was developed was successful in accurately fitting the experimental data for the helium via a best-fit of the friction coefficient or heat production at the interface of the rubbing samples.

The success of the model for the simplest case (i.e., a low pressure non-reactive gas) led to the hope that similar results could be achieved in much more difficult situations (i.e., a high-pressure reactive gas). The goal of this project was therefore to expand the low pressure model to include high pressure oxygen and nitrogen systems (at least up to 1,000 psia).

MODEL DEVELOPMENT

The model used in this study is the same transient one-dimensional equation for heat conduction as in the original model.[2] For each annular sample (i = 1,2) we have the following energy balance

$$\rho_i C_{p,i} \frac{\partial T}{\partial t} = \frac{\partial}{\partial x} k_i \frac{\partial T}{\partial x} - Q_v - Q_c - Q_r \qquad (1)$$

where ρ_i = the density of sample i, $C_{p,i}$ = the heat capacity of sample i, T = temperature in the sample, t = time, x = axial position in the sample, k_i = the thermal conductivity of sample i, Q_v = the heat loss per unit rod volume due to convection, Q_c = the heat loss per unit rod volume due to conduction and Q_r = the heat loss per unit rod volume due to radiation.

The initial condition is:

$$T = T_o \quad \text{for } -L_2 \le x \le L_1 \qquad (2)$$

where L_i = the length of the each sample (i = 1 is the stationary sample and i = 2 is the rotating sample).

The boundary conditions are:

At $x = 0$:

$$-k_1 \frac{\partial T}{\partial x} + k_2 \frac{\partial T}{\partial x} = Q_F \qquad \text{and} \qquad T_{0+} = T_{0-} \tag{3}$$

where Q_F is the total energy flux at the interface via friction and/or reaction.

At $x = L_i$ $(i = 1,2)$:

$$T = T_w = T_o \tag{4}$$

Energy Flux at the Interface:

In a nonreactive environment the energy flux Q_F is given by the equation

$$Q_F(t) = P(t)v(t)\mu(t) \qquad \text{in Btu/in}^2 \text{ sec.} \tag{5}$$

The applied pressure (P) and velocity (v) terms are measured quantities but the coefficient of friction (μ) must be determined via a best-fit of the experimental data. For cases where both friction and reaction are important it is possible to simply use the heat flux Q_F as the parameter to be fit. Three correlations are possible for fitting μ or Q_F vs. time:

1. The first model simply states that the parameter to be fit, $p(t)$ (e.g., the heat flux Q_F or friction coefficient μ), will be held constant over a specified time interval Δt_i, i.e.

$$p(t) = p_i \quad \text{for } t_i \le t \le t_i+\Delta t_i \tag{6}$$

2. The second model uses an Nth-degree polynomial to approximate the parameter $p(t)$ via N+1 coefficients, i.e.,

$$p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_N t^N \tag{7}$$

3. The third model uses parameters similar to Model 1 but also uses quadratic interpolation to approximate the value of the parameter $p(t)$ as a continuous function of time for all times $t_i \le t \le t_i+\Delta t_i$. The parameter values used for interpolation are p_i through p_{i+2}.

Gas-Phase Heat Balance:

For the gas phase the following energy balance was used (assuming the gas phase to be well-mixed):

$$m_g C_{v,g} \frac{dT_g}{dt} = 2\pi R_o \int_{-L_2}^{L_1} h_v (T - T_g)\, dx - A_g h_v (T_g - T_w) \qquad (8)$$

where m_g = the mass of gas in the system, $C_{v,g}$ = the constant volume heat capacity of the gas, h_v = convective heat transfer coefficient, T = the rod temperature (varies with position), T_g = the temperature of the gas, T_w = the chamber wall temperature, R_o = the outer rod radius and A_g = the area for heat transfer from the gas to the surrounding chamber.

Due to the high pressures (above 1,000 psia) used for the oxygen and nitrogen systems, the assumption of ideal gas was inappropriate. This was especially true for physical properties such as heat capacity, thermal conductivity and viscosity. It was therefore decided to use the Lee-Kesler modification of the Benedict-Webb-Rubin equation of state[3] to approximate the departure from the ideal gas properties. (The derivation of the departure equations was carried out by a graduate student (K. Bhattacharrya) under NASA Grant NAG9-557 and will be discussed in more detail in his thesis. It is important to note that the equations for the departure function for gas heat capacity given by both reference 3 and the original paper of Lee and Kesler are incorrect.)

External Heat Transfer From Rods:

The following equations are used for the heat transfer losses from the rods:

Convection: The heat loss from the rods due to convection can be described by the equation

$$Q_v = S_v h_w (T - T_g) \qquad (9)$$

where S_v = the surface area per unit rod volume available for convective heat transfer (= $2R_o/(R_o^2-R_i^2)$) and R_i = the inner radius of the annulus. The convective heat transfer coefficient is assumed to be given by a Nusselt number correlation as:

$$Nu = K (Ta/F_g)^a \, Pr^b \qquad (10)$$

where K, a and b are constants given by one of three methods: 1) the correlation of Becker and Kaye[4], 2) the correlation of Eisenberg et al.[5] or 3) a best-fit of the data. The modified Taylor number is defined as

$$(Ta/F_g) \equiv \frac{\omega^2 R_m d^3}{\nu^2 F_g} \tag{11}$$

where ω = the angular velocity of the rotating cylinder, $r_m = (R_c + R_o)/2$, $d = R_c - R_o$, R_c = the radius of the chamber, ν = the kinematic viscosity of the gas and F_g = a shape factor given by Becker and Kaye[4].

For heat conduction <u>from</u> the rods to the gas the Nusselt number is defined as

$$Nu \equiv \frac{2h_v R_o}{k_g} \tag{12}$$

where k_g = the thermal conductivity of the gas. For heat conduction <u>from</u> the gas to the chamber the Nusselt number is defined as

$$Nu \equiv \frac{2h_v R_c}{k_g} \tag{13}$$

Conduction: For the heat transfer from the cylinders to the holders the heat loss due to conduction is approximated by a "pseudo" heat transfer coefficient h_c

$$Q_c = S_{c,i} h_c (T - T_w) = \frac{S_{c,i} k_g}{\Delta x} (T - T_w) \tag{14}$$

where $S_{c,i}$ = the surface area per unit volume of rod i available for conductive heat transfer and Δx = the distance separating the cylinder and holder.

Radiation: The heat loss due to radiative heat transfer is calculated via

$$Q_r = S_{r,i} \sigma F_{i-j} (T^4 - T_w^4) \tag{15}$$

where all temperatures are absolute (°R), σ = Stefan's constant, $S_{r,i}$ = the surface area per unit volume of rod i available for radiative heat transfer and F_{i-j} = the

shape factor for radiative heat transfer from body i to body j.

NUMERICAL SOLUTION

The addition of the physical property package for the gases at high pressures resulted in a significant increase in the computation time for the model. It therefore became apparent that a more efficient method for the spatial discretization as well as the time integration would be required to keep the computation times to realistic values.

For the time integration the original subroutine LSODE was replaced by LSODES (which consists of the same integration methods but was designed to more efficiently handle the sparse Jacobian matrices which result from reducing partial differential equations to a series of ordinary differential equations via spatial discretization). LSODES was written by A.C. Hindmarsh and A.H. Sherman. The Runge-Kutta-Fehlberg program RKF45 of H.A. Watts and L.F. Shampine was also included in the program because of its increased robustness. (Both LSODES and RKF45 were obtained through the NETLIB distribution system.)

For the spatial discretization two different methods were incorporated. One method is the same finite difference method as in the original model but, in order to determine the temperatures at the exact axial thermocouple locations, quadratic interpolation was used when these positions were located between grid points. This allowed a significant reduction in the required number of grid points (from 18 to 5) to achieve the same accuracy with respect to measured vs. calculated temperatures.

The second method of spatial discretization was to use orthogonal collocation on finite elements.[6] This method consists of replacing the original partial differential equation with the ordinary differential equation

At $x = x_n$:

$$\frac{dT_n}{dt} = \frac{1}{h^2} \sum_{i=1}^{N+2} A_{ni} k(T_i) \sum_{j=1}^{N+2} A_{ij} T_j - Q_v(T_n) - Q_c(T_n) - Q_r(T_n) \quad (16)$$

where x_n = the nth collocation point in the element, h = the size of the element, A_{ij} = the elements of the $(N+2) \times (N+2)$

matrix approximating the first derivative via orthogonal collocation[6] and N = the number of interior collocation points in each element. Finite elements were used to simplify the integration of the convective heat transfer contribution in the gas energy balance (which was accomplished via quadrature).

Two elements were used in each rod with the boundary of each element being the point where the sample holders ended. The solution at the collocation point at the boundary between the two elements in each sample was handled in the usual way by setting the axial gradients equal to each other. If we number the four elements sequentially with element 1 being the section of the stationary sample in the holder we get the following equation for the boundary of the elements in the stationary sample

$$\frac{1}{L_i - X_e} \sum_{k=1}^{N^{(1)}+2} A_{1,k} T_k^{(1)} = \frac{1}{X_e} \sum_{k=1}^{N^{(2)}+2} A_{N^{(2)}+2,k} T_k^{(2)}$$

where $x_{e,s}$ = the axial position for the location of the boundary between the elements in the stationary sample, $T_i^{(j)}$ = the temperature in element j next at the collocation point i in each element. For the rotating sample we simply replace the superscript (1) with (4) and (2) with (3) and $x_{e,s}$ with $x_{e,r}$, the axial position for the element boundary in the rotating sample.

The boundary condition at the interface was handled in a similar manner

$$-\frac{1}{X_{e,s}} \sum_{k=1}^{N^{(2)}+2} A_{1,k} T_k^{(2)} + \frac{1}{X_{e,r}} \sum_{k=1}^{N+2} A_{1,k} T_k^{(3)} = Q_F$$

To make the program more robust two parameter fitting subroutines have been added. One method is a direct search multivariable simplex method developed by Nelder and Mead.[7] The other subroutine is LMDIF which uses a modification of the Levenburg-Marquardt algorithm and was written by B.S. Grabow, K.E. Hillstrom and J.J. More. (LMDIF was obtained through the NETLIB distribution system.)

Results

The computer program (now called *FHP* and installed on the computer system at WSTF) was tested on five high pressure

systems, all with oxygen at approximately 1,000 psia, and gave the results shown in Table 1. The error in Table 1 refers to the root mean square error between the measured and calculated values for both thermocouples in the stationary sample, 0.05 and 0.20 inches from the interface. Tests 830-198 and 830-199 used AISI 304 stainless steel and tests 830-305 through 830-307 used AISI-316 stainless steel. Physical properties of these materials vs. temperature were obtained from Touloukin.[8]

TABLE 1. RESULTS FOR HIGH-PRESSURE OXYGEN TESTS

Test	Average Error (°F)
830-198	±26.4
830-199	±63.6
830-305	±14.3
830-306	±18.8
830-307	±19.4

The temperatures in the samples varied from initial temperatures of approximately 70 °F to final measured temperatures of from 1,000 to 1,400 °F.

Figure 1 gives an example of the results of using Model 3 to approximate the energy flux at the interface for test 830-307. Thermocouple TC-702 is located 0.05 inches from the interface of the rotating and stationary samples and TC-703 is 0.20 inches away. Also shown is the predicted interface temperature. For the AISI 304 tests (830-198 and 830-199) the calculated temperature drop between TC-702 and TC-703 was significantly smaller than the measured value. This was probably due to inaccurate data for the physical properties of the metals.

Figure 2 shows a comparison of the heat flux (in Btu/in^{2}sec) at the interface (via Model 3) vs. time (in seconds) for the three repeated tests 830-305, 830-306 and 830-307.

Conclusions

The program *FHP*, developed as a result of the support over the two past summers by the NASA/ASEE Summer Faculty Fellowship

Program, is successful in accurately determining the heat flux and/or friction coefficients necessary to produce the observed temperatures in the WSTF frictional heating apparatus. The program has enough alternatives with respect to spatial discretization, integration and minimization to make it very robust and sufficiently accurate compared to possible inaccuracies in the experiments. The weak points in the program are:

1. The quantitative effects of convection on the heat loss from the samples. At present a simple correlation from studies on annuli with the inner cylinder rotating is used.
2. The physical properties of some of the metals vs. temperature may be inaccurate.
3. The assumption that the holders are at a constant temperature throughout the simulation.
4. The one-dimensional nature of the model.

At present, only the last problem is actively being investigated. As a part of the research under NASA grant NAG9-557 a computer model is being developed in order to determine the possibility of significant radial gradients through the samples for heat fluxes at the interface which vary radially as well as with time. Once this is known a model can then be developed to include angular variations in the heat flux to determine whether there is a significant difference in temperatures at the same axial position but different angular positions.

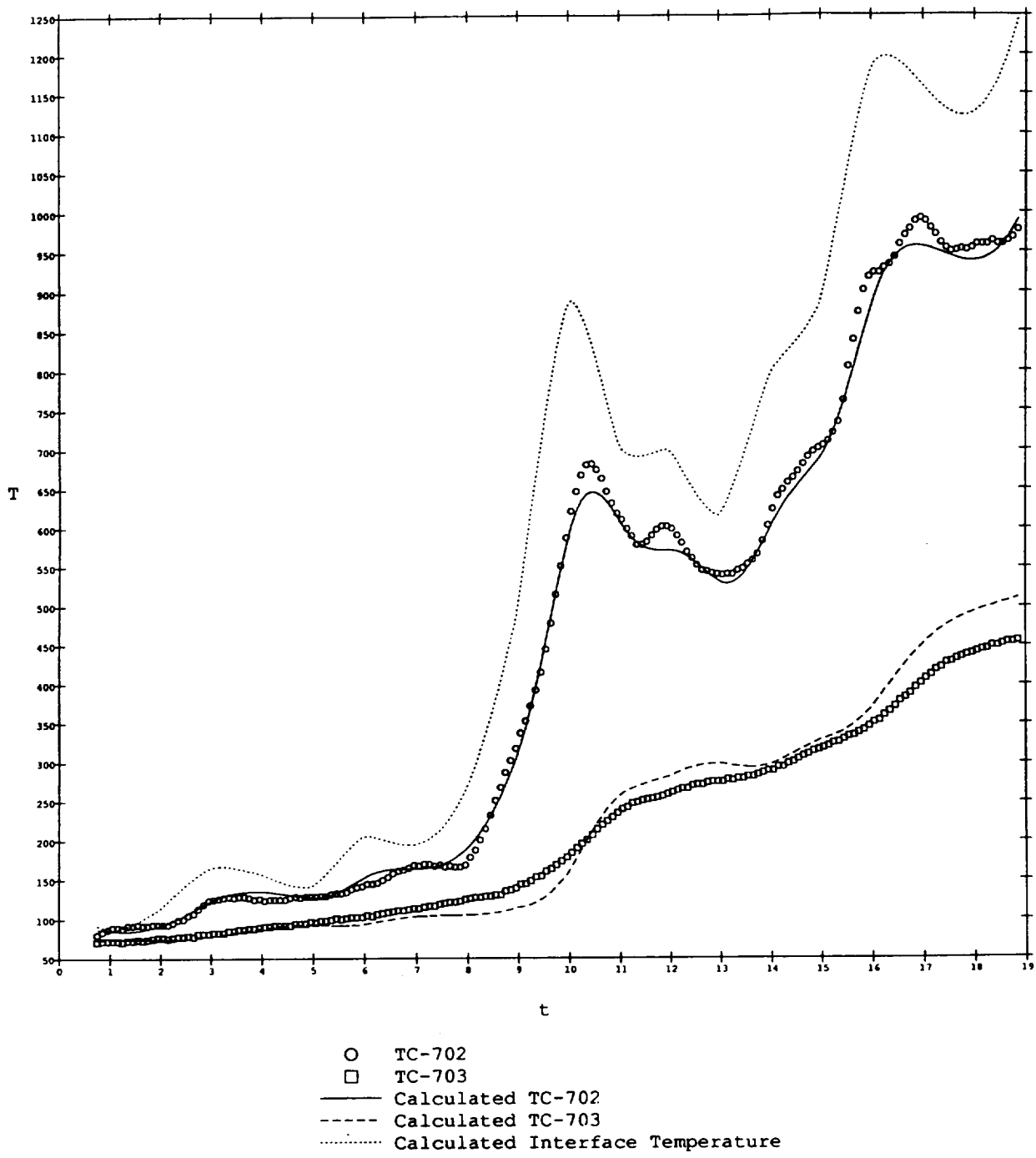


Figure 1. Comparison of measured and calculated temperatures for thermocouples located 0.05 inches (TC-702) and 0.20 inches (TC-703) from the interface. (T in °F and t in sec.)

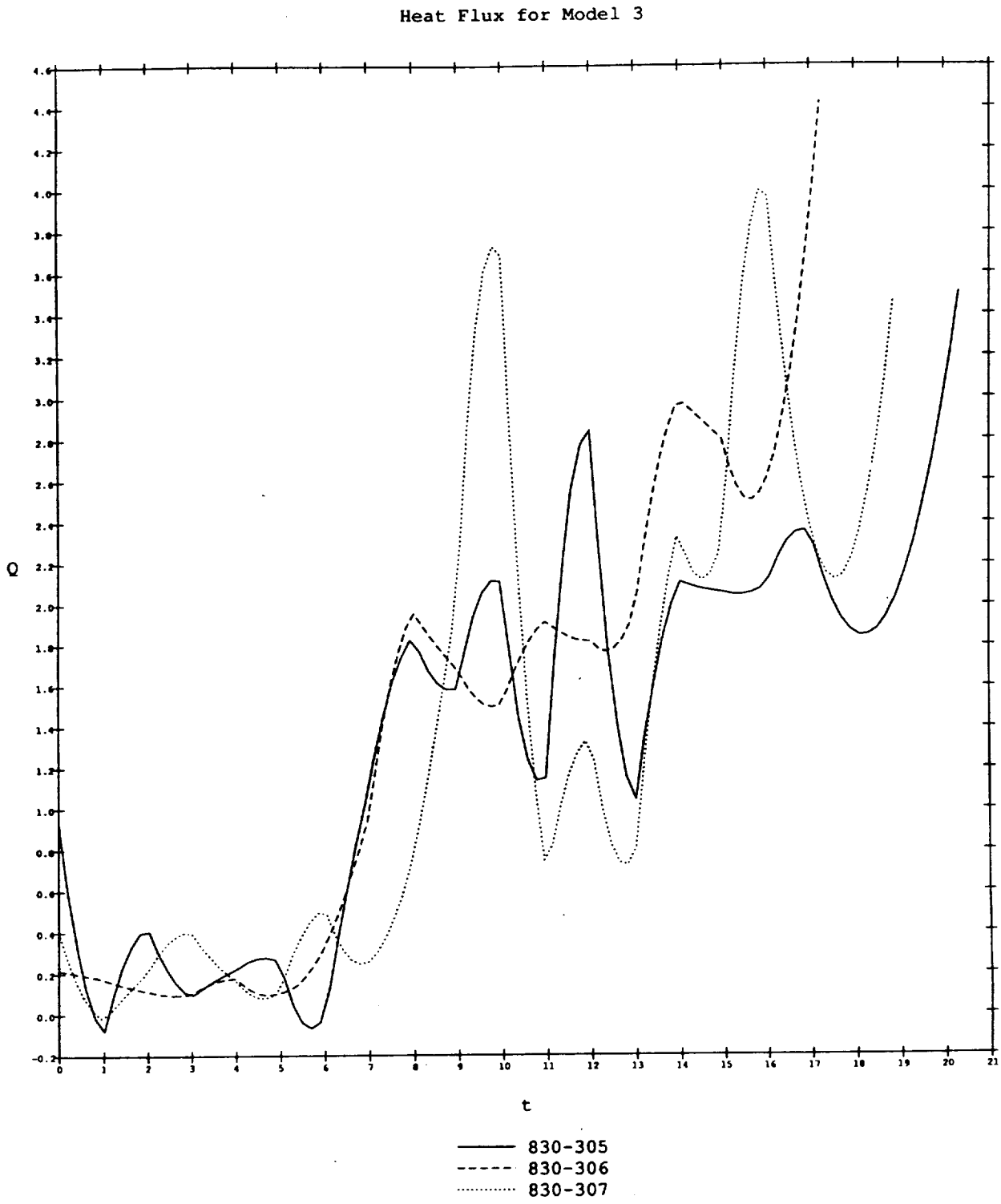


Figure 2. Heat flux at the interface (Q_F in Btu/in^{2}sec.) vs. time (in sec.) for three different tests using Model 3.

REFERENCES

1. Stolzfus, J.M.; Benz, F.J.; and Homa, J.M.: The Pv Product Required for the Frictional Ignition of Alloys. Flammability and Sensitivity of Materials in Oxygen-Enriched Atmospheres: 4th Volume (ASTM STP 1040), J.M. Stolzfus, F.J. Benz and J.S. Stradling, eds., American Society for Testing and Materials (Philadelphia), 1989, pp. 212-223.

2. Skowlund, C.T.; and Beeson, H.D.: Model of the Frictional Heating of Inconel 718 and Titanium (Ti-6V-4Al) in Helium: Final Report for the 1991 NASA/ASEE Summer Faculty Fellowship Program.

3. Reid, R.C.; Prausnitz, J.M.; and Poling, B.E.: The Properties of Gases and Liquids, 4th Edition. McGraw-Hill, 1987.

4. Becker, K.M.; and Kaye, J.: The Influence of a Radial Temperature Gradient on the Instability of Fluid Flow in an Annulus With an Inner Cylinder Rotating. *Transactions of the A.S.M.E., Journal of Heat Transfer*, vol. 84, 1962, pp.106-110.

5. Eisenberg, M.; Tobias, C.W.; and Wilke, C.R.: Mass Transfer at Rotating Cylinders. *Chemical Engineering Progress, Symposium Series No. 16*, vol. 51, 1955, pp.1-16.

6. Finlayson, B.: Nonlinear Analysis in Chemical Engineering. McGraw-Hill, 1980.

7. Nelder, J.A.; and Mead, R.: A Simplex Method for Function Minimization. *Computer Journal*, vol. 7, 1965, pp. 308-313.

8. Touloukin, Y.: Thermophysical Properties of High Temperature Solid Materials: Vol. 3 Ferrous Alloys, The MacMillan Co., 1969.

N93-26080

SPACELAB, SPACEHAB, AND SPACE STATION FREEDOM PAYLOAD INTERFACE PROJECTS

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1992

Johnson Space Center

Prepared By:	Dean Lance Smith, Ph.D.
Academic Rank:	Associate Professor
University & Department:	Memphis State University Department of Elec. Engin. Memphis, Tennessee 38152
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Life Sciences Project
Branch:	Science Operations Branch
JSC Colleague:	Walter Hanby
Date Submitted:	August 7, 1992
Contract Number:	NGT-44-005-803

ABSTRACT

Contributions were made to several projects. Howard Nguyen was assisted in developing the Space Station RPS (Rack Power Supply). The RPS is a computer controlled power supply that helps test equipment used for experiments before the equipment is installed on Space Station Freedom.

Ron Bennett of General Electric Government Services was assisted in the design and analysis of the Standard Interface Rack Controller hardware and software. An analysis was made of the GPIB (General Purpose Interface Bus), looking for any potential problems while transmitting data across the bus, such as the interaction of the bus controller with a data talker and its listeners. An analysis was made of GPIB bus communications in general, including any negative impact the bus may have on transmitting data back to Earth.

A study was made of transmitting digital data back to Earth over a video channel. A report was written about the study and a revised version of the report will be submitted for publication. Work was started on the design of a PC/AT compatible circuit board that will combine digital data with a video signal. Another PC/AT compatible circuit board is being designed to recover the digital data from the video signal. A proposal was submitted to support the continued development of the interface boards after the author returns to Memphis State University in the fall.

A study was also made of storing circuit board design software and data on the hard disk server of a LAN (Local Area Network) that connects several IBM style PCs. A report was written that makes several recommendations.

An preliminary design review was started of the AIVS (Automatic Interface Verification System). The summer was over before any significant contribution could be made to this project.

INTRODUCTION

Contributions were made to several projects as part of the NASA/ASEE Summer Faculty Fellowship Program. Most of the summer was spent on the design of computer interface boards that will permit digital data to be transmitted over a video channel. However, contributions were also made to the SSRPS (Space Station Rack Power System) and the SIRC (Standard Interface Rack Controller). A study was made of using a LAN to support design software. A preliminary design review was started on AIVS (Automatic Interface Verification System). The projects and contributions are briefly described below.

THE SPACE STATION RPS

The SSRPS (Space Station Rack Power System) is a computer controlled power supply that simulates power supply conditions on the Space Station Freedom. The RPS is being developed by Howard Nguyen of the Science Operations Branch (SE2) under the direction of Walter Hanby. The power supply will be used to test equipment that will support experiments on the space station before the equipment is installed in the orbiting laboratory.

The design of the Space Station RPS was reviewed with Mr. Nguyen. The design of the software was reviewed with Sheri Brackett of General Electric Government Services. Several suggestions and alternatives were discussed in several meetings with Mr. Nguyen and one meeting with Ms. Brackett. The design was also reviewed in a meeting with Ken Caldwell, Walter Hanby, Rand Nichols, Mr. Nguyen, and Ms. Brackett.

STANDARD INTERFACE RACK CONTROLLER

Versions of the SIRC (Standard Interface Rack Controller) are currently being used in Spacelab and Spacehab, and a somewhat modified version is expected to be used in the Space Station Freedom. The SIRC provides a standard rack interface to support equipment used for experiments flown on Spacelab, Spacehab, and Space Station Freedom. Each version of the SIRC adds some new features, but tries to maintain compatibility with the earlier versions of the SIRC.

The design of several aspects of the SIRC was reviewed with Ron Bennett of General Electric Government Services who is working with Walter Hanby on the project. The hardware for the SIRC was discussed and several suggestions were made. The possibility of making the serial interface compatible with both RS-232C and RS-422 was discussed.

The design of the system and the software was discussed

with Mr. Bennett, and the software was discussed with David Altmeyer and Robert Deutschman, both of whom also work for General Electric Government Services. Several suggestions were made. The software design was also reviewed at a meeting attended by Walter Hanby and several employees of GEGS.

The SIRC uses a GPIB (General Purpose Interface Bus) for control and low data rate transfers. The GPIB is sometimes referred to by its standard number, IEEE 488. Several issues involving the GPIB were discussed with both Mr. Bennett and Mr. Deutschmann, including how to suspend communication between a talker (data sender in GPIB parlance) and a listener (data receiver in GPIB terminology).

MULTIPLEXING DATA AND A COMPOSITE VIDEO SIGNAL

Most of the summer's effort was spent developing a communication system that will transmit digital data over a composite video signal. The system will be used to increase the digital data communication channel capacity for some of the experiments flown on the Orbiter.

A review was conducted of the pertinent literature. A report was written that describes some of the techniques that have been successfully implemented or proposed. [1] The report makes several design recommendations. A modified version of this report will be submitted to IEEE AES Magazine for publication.

Work was started on specifications for the data communications system. [2] A preliminary design was started on two interface boards for IBM PC/AT or compatible computers. One board will combine the data with the composite video signal. [3] The other board will strip the data from the video signal. [4] A proposal was submitted to support the work after the author returns to Memphis State University in the fall. [5]

CIRCUIT BOARD DESIGN LAN

A study was made of storing circuit board design software and data files on the hard disk server for a LAN that connects several high performance IBM style PCs (Personal Computers). The study was made with the assistance of Hasan Rahman, James Hodges, and Ron Bennett, all of General Electric Government Services. The results of the study are reported elsewhere. [6]

AUTOMATIC INTERFACE VERIFICATION SYSTEM

The design of the AIVS (Automatic Interface Verification

System) was discussed with Ron Bennett of General Electric Government Services. However, no additional work was performed on this project.

REFERENCES

[1] Dean Lance Smith, "A study of transmitting digital data on the Spacelab's analog video channel," to be submitted to *IEEE AES Mag.*

[2] Dean Lance Smith, "Preliminary specifications for the data on composite video system," memo to Walter Hanby, Aug. 7, 1992.

[3] Dean Lance Smith, "Preliminary design of the data on composite video generator board," memo to Walter Hanby, Aug. 7, 1992.

[4] Dean Lance Smith, "Preliminary design of the data on composite video receiver board," memo to Walter Hanby, Aug. 7, 1992.

[5] Dean Lance Smith, "Digital data, composite video multiplexer and demultiplexer boards for an IBM PC/AT compatible computer," Informal Proposal to the Summer Faculty Fellows JSC Director's Grant Program, NASA/JSC, July 22, 1992.

[6] Dean Lance Smith, "Preliminary report on installing CADstar, Maxi/PC, and Maxi/Route on a LAN hard disk server," memo to Walter Hanby, July 31, 1992.

N93-26081

MODEL REDUCTION FOR SPACE STATION FREEDOM

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1992

Johnson Space Center

Prepared By:	Trevor Williams, Ph.D.
Academic Rank:	Associate Professor
University & Department:	University of Cincinnati Department of Aerospace Engineering Cincinnati, Ohio 45221
NASA/JSC	
Directorate:	Engineering
Division:	Guidance, Navigation & Aeronautics
Branch:	Control & Guidance Systems
JSC Colleague:	John Sunkel, Ph.D.
Date Submitted:	August 19, 1992
Contract Number:	NGT-44-005-803

ABSTRACT

Model reduction is an important practical problem in the control of flexible spacecraft, and a considerable amount of work has been carried out on this topic. Two of the best-known methods developed are *modal truncation* and *internal balancing*. Modal truncation is simple to implement but can give poor results when the structure possesses clustered natural frequencies, as often occurs in practice. Balancing avoids this problem but has the disadvantages of high computational cost, possible numerical sensitivity problems, and no physical interpretation for the resulting balanced "modes".

The purpose of this work is to examine the performance of the *subsystem balancing* technique developed by the investigator when tested on a realistic flexible space structure, in this case a model of the Permanently Manned Configuration (PMC) of Space Station Freedom. This method retains the desirable properties of standard balancing while overcoming the three difficulties listed above. It achieves this by first decomposing the structural model into subsystems of highly correlated modes. Each subsystem is approximately uncorrelated from all others, so balancing them separately and then combining yields comparable results to balancing the entire structure directly. The operation count reduction obtained by the new technique is considerable: a factor of roughly r^2 if the system decomposes into r equal subsystems. Numerical accuracy is also improved significantly, as the matrices being operated on are of reduced dimension, and the modes of the reduced-order model now have a clear physical interpretation; they are, to first order, linear combinations of repeated-frequency modes.

INTRODUCTION

Model reduction is a very important practical problem related to the control of flexible space structures (FSS), and a considerable amount of work has been carried out on this topic. Well-known methods include *modal truncation* [1], based either on the natural frequencies of the structure or its modal costs, and *balancing* [2] of the entire structure and then truncation to retain a dominant model for it. An advantage of the balancing approach is that it typically yields a more accurate reduced-order model than does simple modal truncation. This is particularly true when the structure possesses clustered natural frequencies, as is often the case for realistic flexible space structures. However, the disadvantages of balancing are its high computational cost, possible numerical sensitivity problems resulting from the large matrices being operated on, and the difficulty involved in providing a physical interpretation for the resulting balanced "modes".

The purpose of this paper is to investigate the practical performance of the alternative *subsystem balancing* technique when tested on a realistic flexible space structure. This method, introduced in [3][4] and further developed in [5], retains the desirable properties of standard balancing while overcoming the three difficulties listed above. This is achieved by first decomposing the structural model into subsystems of highly correlated modes, based on the *modal correlation coefficients* derived in [4] from the controllability and observability Grammian matrices [6] of the structure. Each subsystem is approximately uncorrelated from all others, so balancing each separately and concatenating the dominant reduced-order models obtained yields roughly the same result as balancing the entire structure directly. The computational cost reduction produced by this block-by-block technique is considerable: an operation count reduction by a factor of roughly $1/r^2$ if the system decomposes into r equal subsystems. The numerical accuracy of the resulting reduced-order model is also improved considerably, as the matrices being operated on are of reduced dimension; this avoids the numerical conditioning problems noted in [8][9] for standard balancing. Furthermore, the modes of the reduced model do now permit a clear physical interpretation. This is a consequence of the fact that each correlated subsystem must necessarily only include modes with close natural frequencies. The balanced modes of each subsystem are therefore, to first order, linear combinations of repeated-frequency modes, and so can themselves be taken as an equally valid set of physical modes. Balancing the entire structure, on the other hand, combines modes of widely differing frequencies, making interpretation difficult.

The results obtained using the software described in this report are for the Permanently Manned Configuration (PMC) of Space Station Freedom. Two different "stick models" [11] for this vehicle were studied, for two choices of solar array and radiator orientations. In both cases, the initial 202-mode flexible body models could be reduced to models with between 20 and 30 modes with very little loss of accuracy.

THEORETICAL BACKGROUND

Consider an n-mode model for the structural dynamics of a modally damped, non-gyroscopic, non-circulatory FSS with m actuators and p sensors, not necessarily collocated. This model can be written in modal form [1] as

$$\ddot{\boldsymbol{\eta}} + diag(2\zeta_i\omega_i)\dot{\boldsymbol{\eta}} + diag(\omega_i^2)\boldsymbol{\eta} = \hat{B}\mathbf{u},$$
$$\mathbf{y} = \hat{C}_r\dot{\boldsymbol{\eta}} + \hat{C}_d\boldsymbol{\eta}, \tag{1}$$

where $\boldsymbol{\eta}$ is the vector of modal coordinates, $\mathbf{u}$ that of applied actuator inputs and $\mathbf{y}$ that of sensor outputs, and ω_i and ζ_i are the natural frequency and damping ratio of the i^{th} mode, respectively. For the typical FSS [7], the $\{\zeta_i\}$ are quite low (e.g. 0.5 %), and the $\{\omega_i\}$ occur in clusters of repeated, or nearly repeated, frequencies as a result of structural symmetry.

Defining the state vector $\mathbf{x} = (\dot{\eta}_1, \omega_1\eta_1, \cdots, \dot{\eta}_n, \omega_n\eta_n)^T$ for this structure yields the state space representation $\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u}$, $\mathbf{y} = C\mathbf{x}$, where $A = blkdiag(A)_i$, $B = (B_1^T, \cdots, B_n^T)^T$ and $C = (C_1, \cdots, C_n)$, with

$$A_i = \begin{pmatrix} -2\zeta_i\omega_i & -\omega_i \\ \omega_i & 0 \end{pmatrix}, B_i = \begin{pmatrix} \mathbf{b}_i \\ 0 \end{pmatrix} \text{and } C_i = \begin{pmatrix} \mathbf{c}_{ri} & \mathbf{c}_{di} / \omega_i \end{pmatrix}; \tag{2}$$

$\mathbf{b}_i$ is the i^{th} row of $\hat{B}$, and $\mathbf{c}_{ri}$ and $\mathbf{c}_{di}$ are the i^{th} columns of $\hat{C}_r$ and $\hat{C}_d$, respectively.

The problem studied here is that of obtaining a reduced-order model

$$\dot{\mathbf{x}}_r = A_r\mathbf{x}_r + B_r\mathbf{u},$$
$$\mathbf{y}_r = C_r\mathbf{x}_r \tag{3}$$

for this structure for which the normalized output error

$$\delta^2 = \int \|\mathbf{y}(t) - \mathbf{y}_r(t)\|_2^2 dt \Big/ \int \|\mathbf{y}(t)\|_2^2 dt \tag{4}$$

is acceptably small. Of course, the size of δ will depend on the order, n_r, chosen for the reduced model. A good model reduction procedure should ideally provide information allowing an intelligent choice for n_r to be made so as to achieve a specified δ value.

Two techniques for model reduction that have been extensively studied are those of modal truncation and internal balancing. The new method implemented in this report, subsystem balancing, can be regarded as an intermediate case between the two established techniques. Model reduction by subsystem balancing proceeds by first dividing the given structure into subsystems of highly correlated modes. Each subsystem is then balanced independently, and a reduced-order model for it generated by deleting all balanced states corresponding to Hankel singular values [2] below some specified threshold. Note that the singular value weighting described in [10] could be applied, if desired, without changing the argument in any way. Similarly, frequency weighting of the Hankel singular values can easily be incorporated to deal with input signals which have a known frequency spectrum. This is actually done in the present application, where the inputs are steps (representing thruster firings) rather than the impulses classically considered in

model reduction problems. The resulting reduced-order subsystem models so obtained are then combined to yield a dominant, approximately balanced, reduced-order model for the full system.

USER INTERFACE

This section describes the user interface to the model reduction package which was developed as part of this contract. This software consists of a library of Matlab m-functions, with mrmain calling all the other functions internally. The package is installed on the Sun SparcStation 2 *deimos* in the Integrated Analysis Laboratory in Building 16, and has also been produced in a Macintosh version. The documentation that follows details the user interface for mrmain; listings of this function, together with the second-level functions it calls, are given as an appendix. All functions have extensive in-line documentation, facilitating future use and/or modification.

Input arguments

om: The natural frequencies (rad/s) of the structure, input as either a row or column vector. Any rigid-body modes must precede the flexible modes and be represented by hard zero frequencies.

phia: The influence matrix, in mass-normalized coordinates, corresponding to the specified actuator locations. If the structure has n modes and m actuators, *phia* will be an (n x m) matrix.

phis: Similar to *phia*, but for sensor stations or positions of outputs of interest (e.g. solar array tips).

User responses

Output the time taken for each step?: The time required for each matrix decomposition, etc., is output to the screen if requested. This allows the progress of the model reduction procedure to be monitored, as well as giving an indication of which steps are the most computationally intensive.

Vectorize? (Faster, but requires more storage): In Matlab, for loops are typically an order of magnitude slower to excute than the equivalent "vectorized" operation. For instance, s=0; for i=1:n, s=s+x(i); end; runs considerably slower than does s=x'*ones(n,1). If vectorization is requested, computation of the system Grammian matrices and correlation coefficients is put into the form of vector-matrix operations rather than loops; this is indeed considerably faster, but requires some additional temporary storage arrays.

Structural damping ratio, % (default is 0.5%): The specified damping ratio is applied uniformly to all flexible modes of the full structural model.

Print frequencies in Hz?: The mean frequency of each subsystem can be output in either rad/s or Hz, as desired.

Desired controllability threshold?: This threshold value is used to determine which modes are correlated in a controllability sense. The system is then broken down into

disjoint sets of modes (subsystems), where modes with a controllability correlation coefficient greater than the specified threshold are deemed to be correlated. Taking a threshold value of 0 implies that all modes are considered correlated, i.e. the method reduces to standard balancing. Conversely, a threshold value of 1 implies that no modes are taken together: this is modal truncation. Intermediate values allow the dimensions of the resulting subsystems to be specified to a large extent; reducing the threshold reduces the number of subsystems, so increasing their dimension.

Desired overall threshold?: This threshold is used in a similar fashion to the controllability threshold, but both controllability and observability are now taken into account. This yields the final subsystem distribution output by the program (in *modemap*) and used to obtain the reduced-order model.

Compare step responses?: If requested, the step responses of the full and reduced-order models are computed, plotted, and the relative differences (i.e. reduced-order model error) output for each input-output channel.

Desired truncation measure?: Two types of measure can be used to define the number of modes retained in the reduced-order model. If a positive integer is input, this is taken to be precisely the desired reduced-order model. On the other hand, if a real number in the interval [0, 1) is input, this is taken to be the desired relative error in the reduced-order model step response, and the model order required to achieve this is computed. (Note that this later option is only an approximation, and should only be used as such.)

Output arguments

am, bm, cm: The reduced-order state-space model obtained.

modemap: This matrix specifies which physical modes are grouped into which subsystems in the decomposition based on overall correlation coefficients. The i^{th} column of *modemap* lists the modes making up the i^{th} of these subsystems.

SUMMARY OF RESULTS

Results will now be provided which illustrate the behavior of the subsystem balancing technique when applied to a structural model [11] of the Permanently Manned Configuration (PMC) of Space Station Freedom. This structure possesses light damping (estimated to be 0.5% of critical), and a large number of closely-spaced vibration modes (202 flexible modes below 10 Hz). Two configurations of the PMC were investigated: in the first, the solar arrays are in the station yz-plane ($\alpha = \beta = 0$) and the radiators in the xy-plane ($\gamma = 0$); in the second, the arrays are in the xy-plane and the radiators in the xz-plane ($\alpha = \gamma = 90^\circ, \beta = 0$). The inputs to these models are the 12 Reaction Control System (RCS) thrusters, i.e. the port/starboard and upper/lower x, y and z jets. The measured outputs are the 3 angular rates sensed by the rate gyros on the station avionics pallet. (The movements at other positions of interest, for instance the solar array tips, could also be considered if desired; the method remains exactly the same.)

A first point to examine is the efficiency of subsystem balancing as compared with that of standard balancing. Matlab function **obalreal** in the Robust Control Toolbox is a reliable

implementation of Moore's balancing algorithm; applying this to the PMC models considered requires approximately 3 hours on a SparcStation 2. By contrast, the subsystem balancing implementation provided by mmain requires approximately 3 *minutes*. Furthermore, the bulk of the operations in subsystem balancing are order(n^2), due in part to the use of closed-form Grammians [6], whereas standard balancing is order(n^3). The efficiency advantages of the new approach will therefore become only more pronounced as larger systems are examined.

The role of the threshold coefficient in determining subsystem dimensions can be seen from the following table. The first column gives various choices for the controllability correlation threshold parameter, and columns 2 and 3 show the resulting maximum subsystem dimensions for the two PMC configurations studied (for input axis port upper x). It can be seen that these dimensions do indeed decrease as the threshold increases, as expected. Also, both systems exhibit broadly similar behavior. It can be noted that the evolution of subsystem dimensions is fairly discontinuous: for instance, large changes occur for thresholds between 0.10 and 0.15, whereas there are hardly any differences between 0.30 and 0.45. A consequence of this is that it is not always possible to find a threshold value which will yield a particular maximum subsystem order. However, it is possible to obtain a good working value which gives a totally acceptable subsystem partition. For the system studied here, a maximum subsystem dimension of about 30 leads to about 36-38 individual subsystems (some of which consist of single modes), a good balance; threshold values giving this distribution were chosen as nominal. Using these thresholds, the original 202-mode flexible models were found to be reducible to models with only 20 to 30 modes without introducing significant errors into the resulting step responses.

TABLE 1. - MAXIMUM SUBSYSTEM DIMENSIONS VERSUS THRESHOLD

Threshold	Max dim, $\alpha = 0^\circ$	Max dim, $\alpha = 90^\circ$	$\alpha=90^\circ$, $\zeta=1\%$
0.00	202	202	202
0.05	165	165	199
0.10	146	131	165
0.15	76	87	131
0.20	55	51	118
0.25	55	31	118
0.30	30	20	87
0.35	30	20	76
0.40	17	20	31
0.45	17	18	31

The fourth column of the table illustrates the effect of damping on model reduction. Damping of flexible structures is a very difficult quantity to model, so there is considerable uncertainty in the damping levels to be chosen for Space Station Freedom. If a value of 1% of critical is used instead of the previous "nominal" lavel of 0.5%, the consequent broadening of the peaks of each mode increases the coupling between modes, so increasing the subsystem dimensions. This does not pose any problem, however; increasing the threshold value to 0.4 will again allow the desired dimensions to be obtained.

REFERENCES

1. R.R. Craig, *Structural Dynamics*, New York: Wiley, 1981.

2. B.C. Moore, 'Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction', *IEEE Transactions on Automatic Control*, Vol. 26, Feb. 1981, pp. 17-32.

3. T.W.C. Williams and W.K Gawronski, 'Model Reduction for Flexible Spacecraft with Clustered Natural Frequencies', invited paper, 3rd NASA/NSF/DoD Workshop on Aerospace Computational Control, Oxnard, CA, Aug. 1989.

4. W.K Gawronski and T.W.C. Williams, 'Model Reduction for Flexible Space Structures', *Journal of Guidance, Control, and Dynamics*, Vol. 14, Jan.-Feb. 1991, pp. 68-76.

5. T.W.C. Williams and M. Mostarshedi, 'Model Reduction Results for Flexible Space Structures', presented at 5th NASA/DoD Control-Structures Interaction Technology Conference, Lake Tahoe, Nevada, Mar. 1992.

6. T.W.C. Williams, 'Closed-Form Grammians and Model Reduction for Flexible Space Structures', *IEEE Transactions on Automatic Control*, Vol. 35, Mar. 1990, pp. 379-382.

7. M.J. Balas, 'Trends in Large Space Structure Control Theory: Fondest Hopes, Wildest Dreams', *IEEE Transactions on Automatic Control*, Vol. 27, 1982, pp. 522-535.

8. M.S. Tombs and I. Postlethwaite, 'Truncated Balanced Realization of a Stable Non-Minimal State-Space System', *International Journal of Control*, Vol. 46, 1987, pp. 1319-1330.

9. M.G. Safanov and R.Y. Chiang, 'A Schur Method for Balanced Model Reduction', *Proc. American Control Conference*, 1988, pp. 1036-1040.

10. R.E. Skelton and P. Kabamba, 'Comments on "Balanced Gains and Their Significance for L^2 Model Reduction"', *IEEE Transactions on Automatic Control*, Vol. 31, Aug. 1986, pp. 796-797.

11. K. Schultz, 'PMC Flex Body Data Transfer', Memo SMD-92-2519, Lockheed Engineering & Sciences Company, Houston, Texas, Apr. 7, 1992.

MATLAB PROGRAM LISTINGS

```
function
[am,bm,cm,modemap]=mrmain(om,phia,phis);
%
% An M-function to perform model reduction based
on
% subsystem balancing for a uniformly-damped
% flexible structure with rate outputs
%
% All other functions in this package are called
% by this main routine.
%
% Arguments:
%
% In:  Frequency vector om (rad/s), with any
%      rigid-body modes (om(i)=0) leading;
%      influence matrices phia (actuators)
%      and phis (sensors), one row per mode
%
% Out: Reduced-order state-space model {am,bm,cm}
%      of flexible-body dynamics;
%      the i-th column of modemap lists those modes
%      making up the i-th unreduced subsystem
%

% Trevor Williams, NASA JSC, August 19, 1992
%
n=max(size(om));  % Number of modes considered
%
%  Strip off any rigid-body modes
%
nflex=sum(sign(om)); nrigid=n-nflex;
om=om(nrigid+1:n);
b=phia(nrigid+1:n,:);
c=phis(nrigid+1:n,:)';
n=nflex;
%
timeout=0;
stime=input('Output the time taken for each step? ',
's');
if stime == 'y' timeout=1; end;  % Time o/p wanted
%
vect=0;
svect=input('Vectorize? (Faster, but requires more
storage) ', 's');
if svect == 'y' vect=1; end;  % Vectorization wanted
%
%  Enter specified damping ratio
%
ze=input('Structural damping ratio, % (default is
0.5%) ');
if isempty(ze) == 1 ze=0.5; end; ze=ze/100;
%
radhz=1;
shz=input('Print frequencies in Hz? ', 's');
if shz == 'y' radhz=1/(2*pi); end;  % Output format
%
%  First compute controllability correlation coeffs
%
t0=clock;
ro=cccalc(om,ze,b,vect);
if timeout ~= 0 t1=etime(clock,t0);
  s=[' Ro-c completed after ', num2str(t1), ' s'];
  disp(' '); disp(s);
  end;
%
% Next, find observability Grammian for entire sys
%
t0=clock;
wo=cfgram(om,ze,c',-1,vect);
if timeout ~= 0 t1=etime(clock,t0);
  s=[' Wo completed after ', num2str(t1), ' s'];
  disp(' '); disp(s);
  end;
%
% Input desired controllability correl'n threshold
%
disp(' ');
disp(' Threshold values should lie between 0 and 1;')
disp(' lower values give fewer (but larger)
subsystems.')
disp(' Enter a negative value when finished.')
%
xdum=1;  % Dummy: gives inelegant indefinite loop
while xdum > 0
  romin=input('Desired controllability threshold? ');
  if romin >= 1 romin=1-eps,end;
  if romin >= 0
%
%  Determine subsystems of correlated modes
%
    t0=clock;
    [isort nsub]=subsys(ro,romin);
    if timeout ~= 0 t1=etime(clock,t0);
      s=[' Internal decomposition took ', num2str(t1), '
s'];
      disp(' '); disp(s);
      end;
    kmax=max(size(nsub));  % Num of subsystems
    s=[' Yields ', int2str(kmax), ' subsystems;
maximum size '];
    s=[s, int2str(max(nsub)), ', minimum ',
int2str(min(nsub))];
    disp(' '); disp(s);
%
  else xdum = -1;
  end;
%
  end;
%
%  Set up index to reorder Wo (agrees with isort)
%
iwsort(2:2:2*n)=2*isort;
iwsort(1:2:2*n-1)=2*isort-ones(1,n);
%
```

```
% Operate on each subsystem in turn
%
i1=1;
nmax=max(nsub); % Greatest subsystem order
wctot=[];
t0=clock;
for k=1:kmax
  nsubk=nsub(k);
  i2=i1+nsubk-1; ivect=isort(i1:i2);
  iwvect=iwsort(2*i1-1:2*i2);
%
% First find its controllability Grammian
%
  wck=cfgram(om(ivect),ze,b(ivect,:),1,vect);
%
% ... then find its singular value decomposition
%
  [uk,sk,vk]=svd(wck);
  wck=uk*sqrt(sk);
%
% Finally, apply to correct row & col blocks of Wo
%
  wo(iwvect,:)=wck'*wo(iwvect,:);
  wo(:,iwvect)=wo(:,iwvect)*wck;
%
  wctot=[wctot [wck; zeros(2*(nmax-
nsubk),2*nsubk)]];
  i1=i2+1;
  end;
if timeout ~= 0 t1=etime(clock,t0);
  s=[' Wbar completed after ', num2str(t1), ' s'];
  disp(' '); disp(s);
  end;
%
% Find matrix of overall correlation coefficients
%
t0=clock;
ro=occalc(wo,vect);
if timeout ~= 0 t1=etime(clock,t0);
  s=[' Ro-o completed after ', num2str(t1), ' s'];
  disp(' '); disp(s);
  end;
%
% Now try various overall threshold values
%
xdum = 1; % Dummy variable again
while xdum > 0
  romin2=input('Desired overall threshold? ');
  if romin2 >= 1 romin2=1-eps,end;
  if romin2 >= 0
%
%   Determine new subsystems of correlated modes
%
    t0=clock;
    [isort2 nsub2]=subsys(ro,romin2);
    if timeout ~= 0 t1=etime(clock,t0);
      s=[' Final decomposition took ', num2str(t1), '
s'];
      disp(' '); disp(s);
      end;
    kmax2=max(size(nsub2)); % Num of subsystems
      s=[' Yields ', int2str(kmax2), ' subsystems;
maximum size '];
      s=[s, int2str(max(nsub2)), ', minimum ',
int2str(min(nsub2))];
      disp(' '); disp(s);
%
  else xdum = -1;
  end;
%
  end;
%
% Now that the final subsystems are defined,
% need to define corresponding Wo ordering
%
iwsort2(2:2:2*n)=2*isort2;
iwsort2(1:2:2*n-1)=2*isort2-ones(1,n);
%
% Find balancing transformation for each subsystem;
% store it and the weighted Hankel singular values
%
i1=1;
nmax=max(nsub2); % Greatest subsystem order
ombar=zeros(1,kmax2);
modemap=zeros(nmax,kmax2);
hsv2t=[];
ttot=[];
t0=clock;
%
for k=1:kmax2
  nsubk=nsub2(k);
  i2=i1+nsubk-1; ivect=isort2(i1:i2);
  iwvect=iwsort2(2*i1-1:2*i2);
  modemap(1:nsubk,k)=ivect'+nrigid*ones(nsubk,1);
%
% Solve this sym eigen-/singular value problem
%
  [tk,hsv2k,vk]=svd(wo(iwvect,iwvect));
  hsv2k=diag(hsv2k);
  [hsv2k ihsv]=sort(-hsv2k); hsv2k=-hsv2k;
  tk=tk(:,ihsv); % Sort in descending order
  ttot=[ttot [tk; zeros(2*(nmax-nsubk),2*nsubk)]];
%
% Perform ad hoc step response frequency weighting
%
  ombar(k)=sum(om(ivect))/nsubk;
  hsv2k=hsv2k/(ombar(k)*ombar(k));
  hsv2t=[hsv2t; hsv2k];
%
  i1=i2+1;
  end;
hsv2sum=sum(hsv2t);
[hsv2s ihsv]=sort(-hsv2t); hsv2s=-hsv2s;
if timeout ~= 0 t1=etime(clock,t0);
  s=[' Hankel singular values took ', num2str(t1), ' s'];
  disp(' '); disp(s);
  end;
%
% Compute the balanced full-order system
%
clear wo; % Need the space for am (2nx2n here!)
t0=clock;
```

```
[am bm cm]=ssmodal(om,ze,b,c);
if timeout ~= 0 t1=etime(clock,t0);
  s=[' Full modal model took ', num2str(t1), ' s'];
  disp(' '); disp(s);
  end;
%
%  Now apply Wc similarity transformation, by blocks
%
t0=clock;
[am bm cm]=blkmult(am,bm,cm,wctot,iwsort,nsub);
if timeout ~= 0 t1=etime(clock,t0);
  s=[' Wc similarity took ', num2str(t1), ' s'];
  disp(' '); disp(s);
  end;
%
%  Then apply T similarity transformation, by blocks
%
t0=clock;
[am bm cm]=blkmult(am,bm,cm,ttot,iwsort2,nsub2);
if timeout ~= 0 t1=etime(clock,t0);
  s=[' T similarity took ', num2str(t1), ' s'];
  disp(' '); disp(s);
  end;
%
%  Set up for step response calculations, if wanted
%
disp(' ');
sstep=input('Compare step responses? ', 's');
if sstep == 'y'
  t=4/(ze*min(om));   % Time for decay
  t=100*round(t/100);  % Round to nearest hundred
  dt=t/100; t=[dt:dt:t];  % 100 points
%
  m=size(b); m=m(2);   % Number of inputs
  p=size(c); p=p(1);   % Number of outputs
%
%  Compute and store step responses of full system
  yf=[];
  t0=clock;
  for iu=1:m
     yf=[yf modstep(om,ze,b,c,iu,t)];
     end;
%
  if timeout ~= 0 t1=etime(clock,t0);
    s=[' Full step response took ', num2str(t1), ' s'];
    disp(' '); disp(s);
    end;
  end;
%
%  Try various different truncation measures
%
disp(' ');
disp(' Enter either the desired number of modes
(integer > 0)')
disp(' or the acceptable approx relative output error (<
1);')
disp(' enter a negative quantity when finished.')
%
xdum = 1;  % Dummy variable again
while xdum > 0
   cutoff=input('Desired truncation measure? ');
   if isempty(cutoff) == 1 cutoff=1; end;  % Safety
   t0=clock;
   if cutoff < 0 xdum = -1;
   else
      if cutoff >= 1 nrom=min(cutoff,n);  % # modes
      else  % Find num modes from desired rel err
         abserr2=cutoff*cutoff*hsv2sum;
         test=0; i=n;
         while test <= abserr2
            test=test+hsv2s(2*i)+hsv2s(2*i-1);
            i=i-1;
            end;
         nrom=i+1;
      end;
%
%     Find number of modes kept from each subsys
%
      hsv2min=hsv2s(2*nrom);
      nsubr=zeros(1,kmax2);
      ioff=0;
      for k=1:kmax2
         for i=1:nsub2(k)
            if abs(hsv2t(2*(ioff+i))) >= hsv2min
               nsubr(k)=nsubr(k)+1;
               end;
            end;
         ioff=ioff+nsub2(k);
         end;
%
%     Finally, find truncation index to give ROM
%
      i1=1;
      iavect=[];
%
      for k=1:kmax2
         i2=i1+2*nsubr(k)-1;
         iavect=[iavect iwsort2(i1:i2)];
         i1=i1+2*nsub2(k);
         end;
%
%  Alter cm so as to give zero steady-state error
%
      delcm=0*cm;
      g=am(iavect,iavect)\bm(iavect,:);
      x=-cm(:,iavect)*g/(g'*g);
      delcm(:,iavect)=x*g';
%
      if timeout ~= 0 t1=etime(clock,t0);
        s=[' Model dimension found after ',
num2str(t1), ' s'];
        disp(' '); disp(s);
        end;
%
%     Output subsystem information
%
      disp(' ');
      disp(' Subsystem   Number   Mean freq');
      s='  dimension  retained  ';
      if shz == 'y' s=[s, ' (Hz)']; else s=[s, '(rad/s)']; end;
      disp(s);
%
```

```
      for k=1:kmax2
        spad=' ';
        if nsub2(k) >= 10 spad=' '; end;
        if nsub2(k) >= 100 spad=''; end;
        s=['    ', spad, int2str(nsub2(k))];
%
        spad=' ';
        if nsubr(k) >= 10 spad=' '; end;
        if nsubr(k) >= 100 spad=''; end;
        s=[s, '        ', spad, int2str(nsubr(k))];
%
        spad='  ';
        prombar=radhz*ombar(k);
        if prombar >= 10 spad='  '; end;
        if prombar >= 100 spad='  '; end;
        if prombar >= 1000 spad=''; end;
        s=[s, '    ', spad, num2str(prombar)];
        disp(s);
        end;
%
%    Compare step responses if requested
%
     if sstep == 'y'
       for iu=1:m
         t0=clock;
         amp=am(iavect,iavect);
         bmp=bm(iavect,:);
         cmp=cm(:,iavect)+delcm(:,iavect);
         yr=blkstep(amp,bmp,cmp,iu,t);
         if timeout ~= 0 t1=etime(clock,t0);
           s=[' ROM step response took ', num2str(t1),
' s'];
           disp(' '); disp(s);
           end;
%
         for io=1:p
           iyf=(iu-1)*p+io;  % Access correct
response
           plot(t,[yf(:,iyf) yr(:,io)]);
           s=['Step responses, nrom = ',
int2str(nrom)];
           s=[s, '; input ', int2str(iu)];
           s=[s,', output ', int2str(io)];
           title(s); grid;
           xlabel('Time (s)');
           ylabel('Outputs');
           pause;
%
           plot(t,(yf(:,iyf)-yr(:,io)));
           s=['Step error, nrom = ', int2str(nrom)];
           s=[s, '; input ', int2str(iu)];
           s=[s, ', output ', int2str(io)];
           title(s); grid;
           xlabel('Time (s)');
           ylabel('Error');
           pause;
           yerr=norm(yf(:,iyf)-
yr(:,io))/norm(yf(:,iyf));
           s=[' 2-norm relative output error of ',
num2str(yerr)];
           disp(' '); disp(s);
         end;
       end;
     end;
   end;
%
% Finally, store the chosen reduced-order model
%
am=am(iavect,iavect);
bm=bm(iavect,:);
cm=cm(:,iavect)+delcm(:,iavect);
#
########################################
#
function ro=cccalc(om,ze,b,vect);
%
% An M-function to construct the controllability
% correlation coefficients of a uniformly-
% damped flexible structure
%
% The flag vect ~= 0 for Matlab-style vectorized
% operations: faster, but requires extra temp store

% Trevor Williams, NASA JSC, August 19, 1992
%
n=max(size(om));  % Number of modes considered
%
if vect == 0
%
% "Standard" loop over matrix locations
%
 ro=eye(n);       % Initialization (saves time)
 betad=zeros(n,1);  %   "     "
%
% First compute contribution from B
%
 for i=1:n
   for j=i+1:n
     ro(i,j)=b(i,:)*b(j,:)';
     end;
   betad(i)=max(b(i,:)*b(i,:)', eps);
   end
%
 for i=1:n
   for j=i+1:n
     ro(i,j)=abs(ro(i,j))/sqrt(betad(i)*betad(j));
     end;
   end
%
% Now add the frequency effects (Frob-norm version)
%
 for i=1:n
   for j=i+1:n
     g=om(j)/om(i);
     temp=8*ze*ze*g;
     num=(g-1)^2*(g^2+1);
     num=sqrt(temp*((g*temp)+num));
     den=(g+1)*((g-1)^2+(temp/2));


       ro(i,j)=(num/den)*ro(i,j);
```

```
        ro(j,i)=ro(i,j);
        end;
      end
else
%
% Calculations in Matlab-style vectorized form:
% first compute contribution from B
%
  ro=b*b';
  ro=abs(ro);
%
% Avoid singularities caused by zero entries in b,c
%
  betad=max(diag(ro), eps*ones(n,1));
  ro=ro-diag(diag(ro)); % Zero all diagonals
  ro=ro+diag(betad);    % Put back diag, or eps
%
  betads=sqrt(diag(betad));
  ro=betads\abs(ro)/betads;
%
% Now add the frequency effects (Frob-norm version)
%
  temp=8*ze*ze;
%
  som=size(om);
  if som(1) > 1   % om entered as a column vector
    g=(om.^(-1))*om';
  else            % om entered as a row vector
    g=(om.^(-1))'*om;
  end
%
  num=((g-ones(n)).^2).*((g.^2)+ones(n));
  num=sqrt((temp*g).*((temp*(g.^2))+num));
  den=(g+ones(n)).*(((g-ones(n)).^2)+(temp*g/2));
  ro=(num./den).*ro;
end
#
##############################################
#
function w=cfgram(om,ze,b,cobs,vect);
%
% An M-function to compute the closed-form
% Grammians of a uniformly-damped flexible
% structure with rate measurements
%
% The flag cobs == 1 for controllability,
%              -1 for observability.
%
% The flag vect ~= 0 for Matlab-style vectorized
% operations: faster, but requires extra temp store

% Trevor Williams, NASA JSC, August 19, 1992
%
n=max(size(om)); % Number of modes considered
w=zeros(2*n); % Initialization (saves time later)
%
if vect == 0
%
% "Standard" loop over matrix locations
%
% Compute each (2 x 2) Grammian block in turn,
% and store in correct upper & lower locations in W
%
for i=1:n
  omi=om(i);
  iw=2*i-1;
  for j=i:n
    omj=om(j);
    jw=2*j-1;
    t1=ze*(omi+omj);
    t2=(omj*omj)-(omi*omi);
    t3=2*omi*omj;
    t4=t3*t1;
    dij=2*t3*(t1*t1)+(t2*t2);
    wij=[t4 cobs*omj*t2; -cobs*omi*t2 t4];
    wij=((b(i,:)*b(j,:)')/dij)*wij;
    w(iw:iw+1,jw:jw+1)=wij;
    w(jw:jw+1,iw:iw+1)=wij';
    end;
  if b(i,:)*b(i,:)' < eps  % Avoid singularity
    w(iw:iw+1,iw:iw+1)=eps*eye(2);
    end;
  end;
else
%
% Calculations in Matlab-style vectorized form:
% first compute contribution from B
%
  i1=(1:2:2*n-1);
  i2=(2:2:2*n);
%
  som=size(om);
  if som(1) > 1   % om entered as a column vector
    t1=ze*(om*ones(1,n)+ones(n,1)*om');
    t2=(ones(n,1)*(om.^2)'-(om.^2)*ones(1,n));
    omj=ones(n,1)*om';
    t3=2*om*om';
  else            % om entered as a row vector
    t1=ze*(om'*ones(1,n)+ones(n,1)*om);
    t2=(ones(n,1)*(om.^2)-(om.^2)'*ones(1,n));
    omj=ones(n,1)*om;
    t3=2*om'*om;
  end
%
  d=2*t3.*(t1.^2)+(t2.^2);
  beta=b*b';
%
% Avoid singularities caused by zero entries in b,c
%
  betad=max(diag(beta), eps*ones(n,1));
  beta=beta-diag(diag(beta)); % Zero all diagonals
  beta=beta+diag(betad);   % Put back diag, or eps
%
  w(i1,i1)=beta.*t3.*t1./d;
  w(i1,i2)=cobs*beta.*omj.*t2./d;
  w(i2,i1)=-cobs*beta.*omj'.*t2./d;
  w(i2,i2)=beta.*t3.*t1./d;
end
#
##############################################
#
function [isort, nsub]=subsys(ro,romin);
```

```
%
% An M-function to determine those modes which are
% deemed to be correlated, based on a given matrix
% of modal correlation coefficients and a
% specified threshold value
%
% Outputs:
%
%  isort: sorting required to produce subsystems
%  nsub:  dimensions of the subsystems

% Trevor Williams, NASA JSC, August 19, 1992
%
n=max(size(ro));  % Number of modes considered
%
% Perform various initializations
%
num=0;      % Total number of modes grouped
nsub=[];    % No subsystem dimensions yet
isort=[1:]; % Modal ordering: unchanged so far
kvect=(n+1)*ones(1,n); % Makes unsorted modes last
kflag=0;  % 1 for subsystem 1, 2 for subsys 2, etc.
%
% Characterize each subsystem in turn
%
while num < n
%
%  Initializations for this subsystem
%
  nsubk=0;            % No modes yet found
  kflag=kflag+1;      % Increment flag
  itest=isort(num+1); % Mode to be tested 1st
  jtest=isort(num+1:n); % Test for corr'n to i
%
  while isempty(itest) == 0
%
%     itest contains a set of modes to be tested
%
    inew=[]; % No modes for next pass yet
%
    for i=itest
      for j=jtest
        if ro(i,j) >= romin & kvect(j) > n
%
%           Mode j is a new mode, correlated
%           to i: store in this subsystem
%
            kvect(j)=kflag;
            nsubk=nsubk+1;
            if j ~= i inew=[inew j]; end;
            end;
          end;
        end;
%
%    Pick up new set of modes (if any) to test
%
    itest=inew;
    end
%
%  This subsystem is finished: store its data
%
  num=num+nsubk;    % Total number of modes
  nsub=[nsub nsubk]; % Subsystem dimensions
  [kvsort isort]=sort(kvect); % Sorted modes
  end
#
#########################################
#
function ro=occalc(w,vect);
%
% An M-function to compute the overall
% correlation coefficients of a uniformly-
% damped flexible structure
%
% The flag vect ~= 0 for Matlab-style vectorized
% operations: faster, but requires extra temp store

% Trevor Williams, NASA JSC, August 19, 1992
%
n=max(size(w))/2; % Number of modes considered
%
if vect == 0
%
% "Standard" loop over matrix locations
%
 ro=eye(n); % Initialization (saves time later)
 for i=1:n
   iw=2*i-1;
   wiif=norm(w(iw:iw+1,iw:iw+1),'fro');
   for j=i+1:n
     jw=2*j-1;
     wjjf=norm(w(jw:jw+1,jw:jw+1),'fro');
     wijf=norm(w(iw:iw+1,jw:jw+1),'fro');
     ro(i,j)=wijf/sqrt(wiif*wjjf);
     ro(j,i)=ro(i,j);
     end;
   end
else
%
% Frobenius norm calculation in Matlab vector form
%
 n2=2*n;
 e=eye(n2)+diag(ones(n2-1,1),-1);
 e=e(:,1:2:n2-1);
%
% Calculate Frobenius norms of each (2x2) block
%
 ro=e'*(w.*w)*e;
 ro=sqrt(ro);
%
% Now normalize
%
 rodiags=diag(sqrt(diag(ro)));
%
 ro=rodiags\ro/rodiags;
end
#
#########################################
#
function
[amodal,bmodal,cmodal]=ssmodal(om,ze,b,c);
%
```

```
% An M-function to construct the modal state
% space model corresponding to a uniformly-
% damped flexible structure

% Trevor Williams, NASA JSC, August 19, 1992
%
n=max(size(om));      % Number of modes considered
%
amodal=zeros(2*n);
adiag=[-2*ze -1;1 0];  % "Template" for diags of a
%
m=size(b); m=m(2);
bmodal=zeros(2*n,m);
%
p=size(c); p=p(1);
cmodal=zeros(p,2*n);
%
for i=1:n
   i2=2*i-1;
   amodal(i2:i2+1,i2:i2+1)=om(i)*adiag;
   bmodal(i2,:)=b(i,:);
   cmodal(:,i2)=c(:,i);
   end
#
##########################################
#
function
[am,bm,cm]=blkmult(am,bm,cm,wctot,iwsort,nsub);
%
% An M-function to apply a similarity
% transformation stored as ordered blocks
% to a given state space model

% Trevor Williams, NASA JSC, August 19, 1992
%
kmax=max(size(nsub));  % Number of subsystems
%
iw1=1;
for k=1:kmax
   nsubk2=2*nsub(k);
   iw2=iw1+nsubk2-1;
   iwvect=iwsort(iw1:iw2);  % Required state order
%
%  Retrieve k-th block of similarity transformation
%
   wck=wctot(1:nsubk2,iw1:iw2);
%
%  Premult row blocks of am, bm by inverse of wck
%
   am(iwvect,:)=wck\am(iwvect,:);
   bm(iwvect,:)=wck\bm(iwvect,:);
%
%  Postmultiply column blocks of am, cm by wck
%
   am(:,iwvect)=am(:,iwvect)*wck;
   cm(:,iwvect)=cm(:,iwvect)*wck;
%
   iw1=iw2+1;
   end
#
##########################################
#
function y=modstep(om,ze,b,c,iu,t);
%
% An M-function to compute the step response
% of a state space model in "symmetric"
% modal form of a flexible structure

% Trevor Williams, NASA JSC, August 19, 1992
%
n=max(size(om));    % Number of modes
tmax=max(size(t));  % Length of time vector
p=size(c); p=p(1);  % Number of outputs
di=zeros(p,1);
%
y=zeros(tmax,p);
adiag=[-2*ze -1; 1 0];
%
for i=1:n
%
%  Set up required submatrices
%
   ai=om(i)*adiag;
   bi=[b(i,iu); 0];
   ci=[c(:,i), 0*c(:,i)];
%
%  Add step response of this mode to total
%
   y=y+step(ai,bi,ci,di,1,t);
   end
#
##########################################
#
function y=blkstep(a,b,c,iu,t);
%
% An M-function to compute the step response
% of a state space model of a flexible structure
% with A block diagonal (modulo ordering!)

% Trevor Williams, NASA JSC, August 19, 1992
%
tmax=max(size(t));    % Length of time vector
p=size(c); p=p(1);    % Number of outputs
di=zeros(p,1);
%
y=zeros(tmax,p);
%
%  Determine OVERALL block structure of a, directly
%
[iasort,nasub]=subsys(abs(a)+abs(a'),eps);
kmax=max(size(nasub));
%
i1=1;
for k=1:kmax
   i2=i1+nasub(k)-1;
   ia=iasort(i1:i2);
%
%  Add step response of this block to total
%
   y=y+step(a(ia,ia),b(ia,iu),c(:,ia),di,1,t);
   i1=i2+1;
   end
```

N93-26082

THE ADULT LITERACY EVALUATOR: AN INTELLIGENT COMPUTER-AIDED TRAINING SYSTEM FOR DIAGNOSING ADULT ILLITERATES

Final Report

NASA/JSC Summer Faculty Fellowship Program--1992

Johnson Space Center

Prepared by:	David B. Yaden, Jr., Ph.D.
Academic Rank:	Associate Professor
University & Department:	University of Houston Department of Curriculum & Instruction Houston, Texas 77204-5872
NASA/JSC	
Directorate:	Information Systems
Division:	Information Technology
Branch:	Software Technology
JSC Colleague:	James A. Villarreal
Date Submitted:	August 3, 1992
Contract No:	NAG 9-562

ABSTRACT

An important part of NASA's mission involves the secondary application of its technologies in the public and private sectors. One current application being developed is the The Adult Literacy Evaluator, a simulation-based diagnostic tool designed to assess the operant literacy abilities of adults having difficulties in learning to read and write. Using ICAT system technology in addition to speech recognition, closed-captioned television (CCTV), live video and other state-of-the art graphics and storage capabilities, this project attempts to overcome the negative effects of adult literacy assessment by allowing the client to interact with a intelligent computer system which simulates real-life literacy activities and materials and which measures literacy performance in the actual context of its use. The specific objectives of the project are as follows: (a) To develop a simulation-based diagnostic tool to assess adults' prior knowledge about reading and writing processes in actual contexts of application, (b) to provide a profile of readers' strengths and weaknesses, and (c) to suggest instructional strategies and materials which can be used as a beginning point for remediation. In the first and developmental phase of the project, descriptions of literacy events and environments are being written and functional literacy documents analyzed for their components. Examples of literacy events and situations being considered included interactions with environmental print (e.g., billboards, street signs, commercial marquees, storefront logos, etc.), functional literacy materials (e.g., newspapers, magazines, telephone books, bills, receipts, etc.) and employment-related communication (i.e., job descriptions, application forms, technical manuals, memorandums, newsletters, etc.). Each of these situations and materials is being analyzed for its literacy requirements in terms of written display (i.e., knowledge of printed forms and conventions), meaning demands (i.e., comprehension and word knowledge) and social situation. From these descriptions, scripts are being generated which define the interaction between the student, an on-screen guide and the simulated literacy environment. The proposed outcome of the Evaluator is a diagnostic profile which will present broad classifications of literacy behaviors across the major areas of metacognitive abilities, word recognition, vocabulary knowledge, comprehension and writing. From these classifications, suggestions for materials and strategies for instruction with which to begin corrective action will be made. The focus of the Literacy Evaluator will be essentially to provide an expert diagnosis and an interpretation of that assessment which then can be used by a human tutor to further design and individualize a remedial program as needed through the use of an authoring system.

Adult Literacy Evaluator: An Intelligent Computer-Aided Training System for Diagnosing Adult Illiterates

I. Introduction

This report describes the current status, milestones achieved and projected schedule of the Adult Literacy Evaluator (ALE). This spin-off project is attempting to integrate expert system, speech analysis and virtual reality technology developed at NASA into the field of adult literacy with the primary goal of creating a tool for literacy assessment with high ecological validity. The outcomes of this assessment tool will also include written diagnostic profiles and exemplar lessons in remediation strategies to be used by volunteers and staff at literacy provider organizations.

The following sections in this report include descriptions of the current project objectives, past and current developments in theory and the expert system knowledge base, the conceptual model, an updated project plan and schedule, past presentations and publications by the project team, external funding efforts and contractual research in speech analysis and recognition to be incorporated into the ALE.

Finally, the scope of this report covers the time period of March 1991 through June 1992 and describes the project's development under the current principal investigator, Dr. David Yaden. Descriptions of the work effort prior to March 1991 are not included in this document.

II. Current Project Objectives

The project objectives as of 1991 have remained constant and are as follows:

1. To develop a simulation-based diagnostic tool to assess adults' prior knowledge about reading and writing processes in the context of their use;

2. To provide interpretive, diagnostic profiles of readers' strengths and weaknesses; and

3. To provide focused suggestions (i.e., exemplar lessons) for instructional strategies and materials which can be used by volunteer tutors as a beginning point for remediation with adults seeking assistance with literacy providers.

To date, most of the work has been concentrated on objective #1 and has included interface prototyping, knowledge-base development, and the creation of literacy scenarios, i.e., real-life activities in which literacy plays a central role. Work on objectives #2 & 3 are in the initial stages.

The fact that technology for continuous speech recognition is not yet available has necessitated that the project move ahead more quickly on other fronts and develop plans for increasing the scope of the diagnostic capabilities. This increased scope and its attendant project plan is described fully in the following section. The separate development of the speech analysis component for the ALE is described in section IX.

A final caveat to the project objectives is the extent to which the technologies of virtual reality can be fully integrated into the project. Although NASA's Software Technology Branch is spearheading research in this area, the technology is still some years away from being easily incorporated into an application such as the ALE. Therefore, in order to create a virtual environment of authentic literacy materials and behavior, the project team is currently experimenting with a variety of technologies such as live and digitized video pictures with the capability of the interface to depict natural movement in each scene.

III. ALE: A Model for Authentic Assessment - Project Development

Authentic, contextualized assessment is a complex issue. Yet several conceptual hurdles have been successfully maneuvered so far by the Project Team. This section will discuss critical break-throughs made in four major components of the ALE: (a) Scenario development, (b) literacy element knowledge base, (c) diagnostic continuums, and (d) comprehensive profile statements.

A. Scenario Development

1. Multi-leveled assessment. The role of the scenarios was to offer a situation in which a client could enact real literacy actions embedded in purposeful print environments. As a diagnostic tool, the ALE must be able to assess a wide variety of literacy competencies. Unlike other assessment tools which measure a very restricted range of ability, the ALE assumes that adult readers' skills and strategies vary a great deal. Thus, the ALE must be a flexible, responsive system. Also, it is important that the environmental task demands quickly adjust to the clients without their awareness due to the stigma attached to

poor literacy performance. The project team tackled this problem by creating the notion of multi-layered scenarios rather than each having a specific skill/strategy level. This means two clients working next to each other could both be in the grocery store scenario, yet one could be interacting with print at a high level of sophistication such as filling out an application for a grocery card and another could be working at a much lower level, for example, recognizing familiar logos on food items. The system will be created so that the clients' response to each action will determine the rest of the scenario. While preventing certain scenarios from being labeled as "low," this also means that frustration levels can be kept to a minimum (which is a critical issue with adults who have failed academically) and as clients gain new strategies and skills, they will encounter new print challenges within a previously experienced scenario.

2. After-scenarios. Another obstacle emerged within the scenario component. While much can be inferred from a client's interactions with print, there are large arenas of literacy knowledge, critical to traditional analysis, which could not be tapped in the scenario as originally envisioned since speech analysis is not yet available. The project team wrestled with ways to assess more realms of literacy knowledge without speech analysis and recognition. As a solution to this dilemma, the idea of *after-scenarios* was developed. Following each series of contextualized actions, such as renting a movie in a video store, In a separate simulation, the client will be asked more questions about the print experienced in that environment. In this clinical-type interview, the ALE might prompt a client to summarize the blurb on the back of a video or compare three signs in the store and determine which is the most informative. If clients want to see the document back in the original context, they just click on a certain place on the screen and the document reappears as it did in the scenario. The scenario and after-scenario partnership allows for the authenticity of real-life literacy events and the indepth probe of a clinical interview. Currently there is no other assessment tool of which the project team is aware in the education field which combines these two features.

B. Development of the knowledge-base

1. Grounding theory through field study. The field of literacy has yet to agree upon a definition of literacy let alone upon the range of literacy elements involved in making literate decisions. So Dr. Yaden thoroughly reviewed the literature in the field extracting individual literacy elements as they were described. These were then organized logically under traditional key features such as comprehension and fluency, to offer a comprehensive pool of literacy components no where else previously compiled. The next step was to reconnect

the elements pragmatically, in other words, reorganizing the elements so as to reflect a model of literacy which not only includes the components of reading and writing but also the grounded relationships between the components. After reviewing case studies and other work in the field, Drs. Yaden and Brown determined that the information they were seeking was not in the literature and thus it is deemed necessary to undertake a study to further document adult illiteracy. The current plan is to create a grounded theory, that is, a theory of adult literacy that is actually based on observations and interactions with disabled readers in the environment. This grounding will ensure the quality and marketability of the ALE as a revolutionary diagnostic tool.

2. Client informants. The project team is also proposing an innovative design for software development by creating a team of client informants to work as fellow researchers to create the literacy model, scenarios, after-scenarios, and the interface. Rather than building a system and then testing it, the team has decided to work with the informants during the development process. This is critical with a population such as disabled readers who for various circumstances often lead very different lives than the project team members. Scenarios and after-scenarios will be created with the disabled readers' input from start to finish, thus avoiding development of inappropriate tasks. The transition to field testing will be easier due to participation and advice from disabled readers. Again, this is another consideration undertaken by the project team that will impress a business partner.

C. Diagnostic Continuums

1. Expanded scope of literacy assessment. Traditional survey and diagnostic tests in literacy generally measure students' skills in three areas: vocabulary, comprehension, and word recognition. Other facets of reading and writing processes such as metacognition, use of strategies, and literacy concepts are rarely even acknowledged. Since adult illiterates' strengths are often in these unassessed domains, the project team faced another hurdle. The few unconventional tests that do attempt to get at these issues target very young children or are based on oral reading (as mentioned before a highly sophisticated speech analysis system would be necessary for this). The project team approached this problem by utilizing Dr. Brown's previous work which identified expanded typologies of literacy performance and continuums of behavior as opposed to discrete skill levels only. In this earlier study, she documented different dimensions of children's literacy behavior based on observational data and product analysis. Thus after considering this work, Drs. Yaden and Brown decided to create a multi-dimensional profile represented by a series of continuums that will reflect not only traditional skills such as vocabulary and

comprehension, but also reader strategies and concepts of reading and writing. The ALE will break new ground in adult literacy diagnoses with a profile that both details outcome (i.e., performance level) and actions (i.e., strategies and their accompanying knowledge base). Another unusual facet of the ALE continuums is that they will be created through field work with disabled readers. By offering a multi-dimensional profile that reflects critical facets of adult literacy, the ALE promotes a model of remediation which is based on growth and assumes the importance of the socio-psycholinguistic aspects of literacy.

D. Comprehensive Profile Statements

1. Summary of diagnostic interpretations and continuums. In addition to the diagnostic statements interpreting the client's behavior on each of the continuums of literacy performance, the project team has developed the notion of comprehensive profile statements or summaries which will be available to give an overview of the adult's performance across all dimensions of literate behavior. Just as the diagnostic continuums summarize across individual literacy elements, so do the comprehensive statements give a perspective within which all aspects of the adult readers' performance on the ALE can be evaluated. The comprehensive summaries will integrate the adult client's behavior across, for example, performance on textual material, strategies for approaching literate tasks, and concepts about reading and writing events. Therefore, the exemplar lessons supplied to volunteer tutors will be based upon a much broader assessment of the adult's strengths and weaknesses than any current assessment available.

IV. Conceptual Model of the Adult Literacy Evaluator

Figure 1 depicts the current theoretical model of the ALE. The client will go through a series of actions in various literacy scenarios.
Each scenario will involve many natural literacy interactions based on authentic and common life activities, such as going shopping or renting a video.
Monitoring each action of the client, the system will adjust the following literacy demands appropriately. Then an after-scenario will utilize the print materials from that environment to allow for a clinical interview. By presenting specific questions about reading and writing, this component will reveal domains of literacy that would otherwise not be assessed during authentic actions.

Each behavior will be analyzed by an inferencing algorithm, possibly fuzzy logic. Actions will be related to identified literacy elements. Another inferencing system will distribute the elements to diagnositic continuums. These continuums will be based on several broad literacy domains such as print interaction and print-coping strategies. A final inferencing algorithm will analyze each continuum and connect these to one comprehensive profile statement

which will, in turn, be tied to a set of exemplar lessons.

V. Project Plan and ALE Roadmap

The current project plan is represented in Figure 2 and described in detail in the following narrative.

The ALE road map is split into four regions: (a) Field Research refers to work that the literacy team will do in the field to gather in-depth data about the processes and conceptions of adult illiterates; (b) Theory and Literacy Model Development describes the theoretical and conceptual milestones; (c) Software Development details the programming demands; and (d) Grants and Quest for Business Partner relays the on-going search for external funding and a business that will want to market and sell the ALE at the completion of this project. Academic seasons are marked along the road map. Each season will be described below.

Summer 1992

Several fronts will move along this summer. The appropriate avenues at the university will be taken to gain consent for the field work to insure readiness for the fall. Research will continue on the diagnostic continuums and an after-scenario will be developed. The project team will work together to explore new interface possibilities and create a field prototype by summer's end. Writing a R & D grant for the National Literacy Institute will be a time-demanding task which will occupy much of July and early August. This proposal is due August 17, 1992. Efforts will continue to work with Joe Agosta, Director of Development at UH to find external funding.

Fall 92/ Spring 93

During this phase of the project, Drs. Yaden and Brown will conduct a research study on adult disabled readers. During the month of September fifteen tutoring sessions will be observed and the tutor and clients in each partnership will be interviewed. From this broad group, six pairs will be chosen for intensive study over the next eight months. Through observing tutoring sessions and spending time with the disabled readers as they go about their day-to-day activities, a grounded knowledge base will develop. Drs. Yaden and Brown will also begin tutoring their own clients as to have an inside perspective on the process of remediation. During the field work, preliminary development on the scenarios, after-scenarios, exemplar lessons, and inferencing algorithm will continue.

Data analysis, occurring throughout the study, will reveal critical issues in remediation which have not been documented previously. This field work will insure the marketability of the ALE by allowing for the literacy team to create a product grounded in the needs and sensitive to the environmental constraints of adult remediation. Interface testing will enable the refinement of the prototype so that a Field Prototype 2 can be created as the field work is ensuing. It is critical to note that software development will be guided and continually refined by the data from the field. The search for external funding will continue as will the exploration of possible business partners.

Summer 93

From the field research, a few disabled readers will be selected to form an informant group who will work intensely with the research team on several fronts: (a) completing the data analysis; (b) specifying interrelationships of literacy elements; (c) assessing and refining the Field Prototype 2; (d) finalizing the simulation; and (e) finalizing the continuums for the diagnostic profile. With disabled readers as essentially co-researchers immediate feedback will be available and misfires prevented. Components of the ALE will be finalized with the clients present; while being time effective, this should also be a selling point with a business partner.

Fall 93

During this phase the programmers will utilize the information gathered during the previous year to implement the simulation. The literacy team will design detailed scenarios, finalize the simulation to the continuum algorithm and finalize the diagnostic profiles. Hopefully, at this point a business partner will have been secured, if not, the search will intensify.

Spring/ Summer 94

While the literacy team is finalizing the inferencing algorithm from the continuums to the diagnostic profiles, writing exemplar lessons (potentially sixty), and completing each of the after-scenarios, the programmers will be implementing the rest of the ALE components.

Fall 94

This final phase will involve testing in the field, further refinement, and writing of the documentation. The business partner will be integral to this phase as hopefully they would supply monies for field testing. This phase will conclude with a software program ready to be shrink-wrapped and handed off to a business to market and produce.

VI. Presentations

A. Justice Department, JSC.

On March 12, 1991, a presentation of the project was given to a group of Justice Department officials by James Villarreal, Bob Way and David Yaden at the Johnson Space Center. In this presentation, the project's focus as an assessment and diagnostic tool incorporating the concepts of virtual reality was highlighted in comparison to an earlier conception of the system as being a tool for assisting the correct pronunciation of words. At that meeting, the first mock-up demonstration of the interface was presented which included live video, closed-caption TV and digitized pictures. A demonstration of the current speech recognition system being developed by Speech Systems Incorporated was given as well. The redefinition of the project as an assessment tool incorporating NASA's ICAT technology and speech recognition was well-received by those in attendance, and several justice department members volunteered federal prison sites for field-testing.

B. Adult Literacy and Technology Conference, Costa Mesa, CA.

On July 17-20, 1991, David Yaden and Bob Way made a presentation at the Adult Literacy and Technology Conference in Costa Mesa, California. This presentation included the rationale for and description of the simulation-based conception of the project as well as technical information about the research and development platform and the projected final computing requirements. From observation of other CAI programs on display at the conference, it was observed that the ALE is unique in its attempt to assess literacy in context and for the provision of intelligent feedback to both the client and literacy tutor. These features prompted several inquiries about the project both during and after the conference.

C. Representatives from state correctional institutions, JSC.

On November, 1991 at JSC, the project was presented in detail to representatives of the National Institute of Corrections, the NASA Headquarters Technology Utilization Office, the Iowa Department of Corrections, Arizona State Department of Education and Garcz, McGing and Associates. At this meeting both Iowa and Arizona expressed a high degree of interest in the project because of its plan to incorporate real-life literacy skills as the focus of the assessment and for its provision of diagnostic interpretations and suggestions for remediation. Both states agreed to provide prison field-testing sites.

At this meeting as well, the scope of the ALE was outlined and understood to be at least three years away from completion or at a point where it could be handed off to a commercial company for marketing. Primary in this projected timeline was the fact that speech recognition technology is unable at present to analyze continuous speech, a feature necessary for the oral reading analysis component of the ALE. The meeting participants agreed to collaborate on upcoming grants from the Department of Education (DOE) depending upon the status of the ALE at the time funding was available.

VII. Publications

Listed below in order of most recent publication date are articles disseminated by the project team describing ALE. These sources include a newspaper, conference proceedings, peer-reviewed journal and professional organization newsletter.

Yaden, D. B., Brown, L. M., Way, R., & Villarreal, J. A. (in press). Authentic Literacy Assessment: NASA Technology Addressing Adult Illiteracy. *Computers in the Schools.*

Yaden, D. B., & Brown, L. M. (1992). The Adult Literacy Evaluator: NASA's Response to the Literacy Crisis. *Mosaic: Research Notes on Literacy, 2 (2),* 1,4.

Villarreal, J. A., & Yaden, D. B. (1992). A NASA Spin-off: The Adult Literacy Tutor. *Proceedings of the Adult Literacy and Technology Conference.*

Parker, F. (1992, March). Education Professor Using Today's Technology to Help Fight Adult Illiteracy. *UHOUSTON,* March 6, p. 4.

VIII. External Funding

To date, 30 federal and private funding agencies have been contacted by phone or letter describing the ALE and inquiring of their interest in financial support of the project. From these initial inquiries, 14 proposals have been sent to agencies interested and five (Kenan, Coca Cola, Coors, Casey & Beatrice) are still in various stages of review, according to spokespersons from those agencies. The project team is also working closely with the University of Houston Development Office and the Research Triangle Institute in following up on current proposals as well as in exploring new sources of funding not yet contacted.

Department of Education funding 92-93 for corrections, originally anticipated, has not been applied for since the ALE will not ready as a demonstration project by September 30, 1992 when the funds from these grants will be disbursed. However, federal research and development monies for adult education and technology from the National Literacy Institute are available now for competitive applications (due August 14) and a major effort will be expended by the project team in preparing a proposal.

IX. Speech Recognition Research

In the ALE's original conception, speech analysis and recognition was to be one of the primary features of the system. To that end, two efforts have been underway to create an acceptable speech analysis component in order for the ALE to conduct an oral reading analysis, a fundamental aspect of any reading diagnosis.

A. Speech Systems Incorporated.

In the Spring of 1991, SSI delivered to JSC a preliminary version of the demonstration software with peripheral hardware to be run on a Sun platform. This system using a combination of scoring normalization techniques such as syntactic error modeling, score normalization and phoneme error modeling, resulted in a system which had an overall 80 % accuracy rate for target words, an 80% accuracy rate for rejecting non-targets, but less than a 5% percent accuracy rate for recognizing all of the phonemes in a given word. Higher accuracy rates for this system are based on the recognition of one-syllable words, for recognizing all of the phonemes in a multi-syllable word, the system's accuracy is less that 3%.

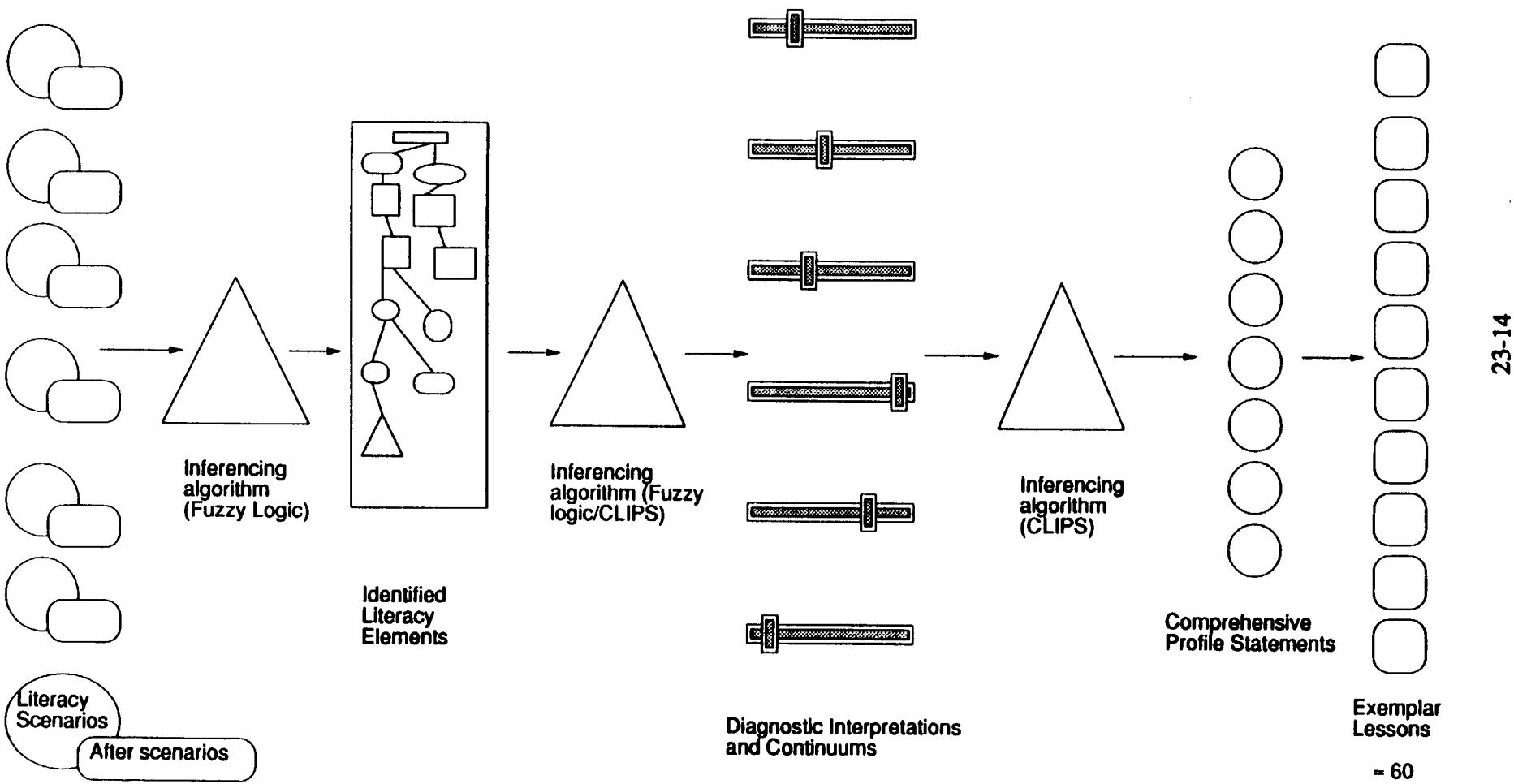
Several aspects of the above system make it difficult to incorporate into the current project: (a) The cost of the Sun platform and necessary peripherals, (b) the fact that ALE is being designed on the Macintosh, and (c) the extremely low accuracy rates for determining all of the phonemes in both one-syllable and multi-syllable words.

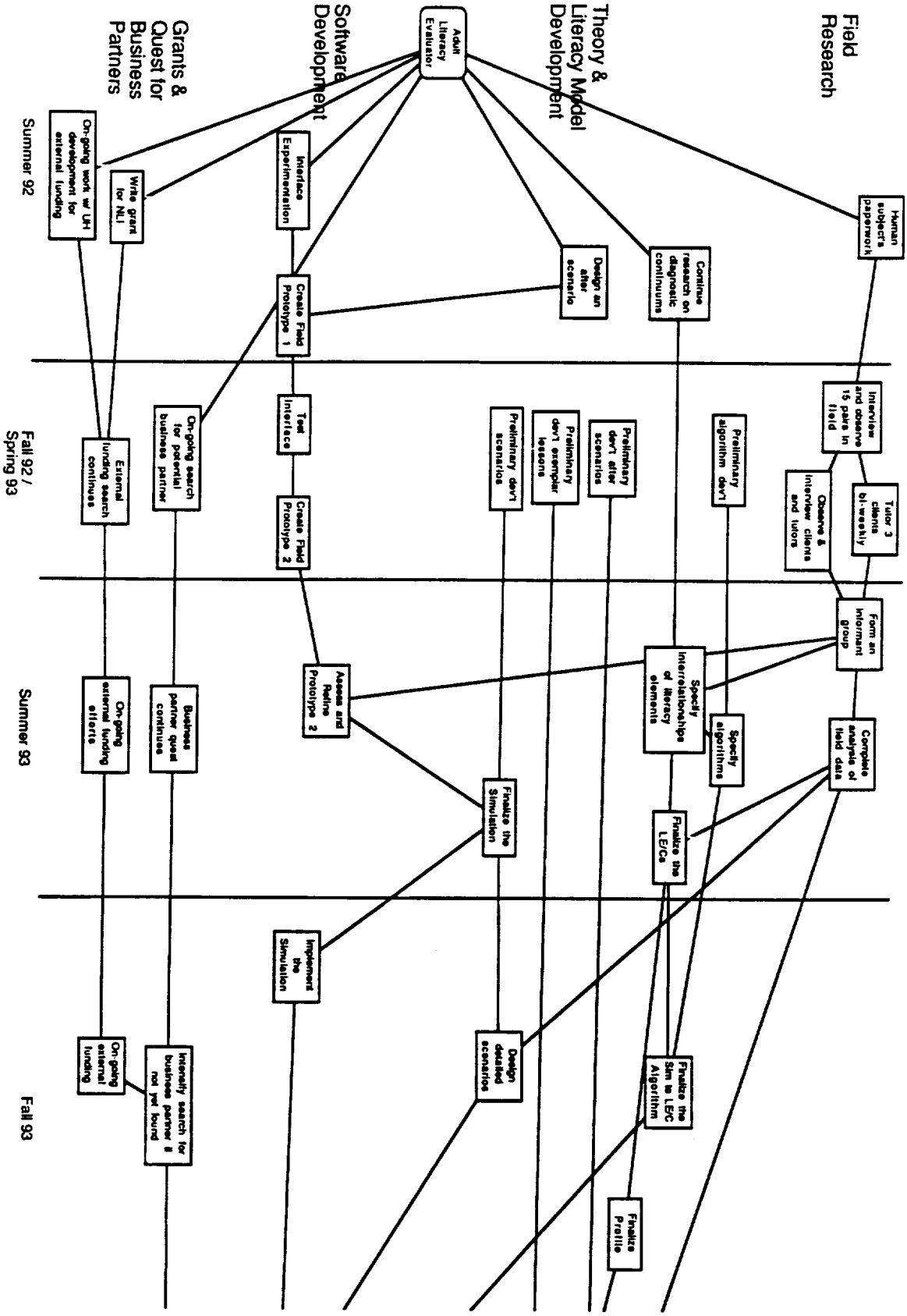
B. DKR Consulting.

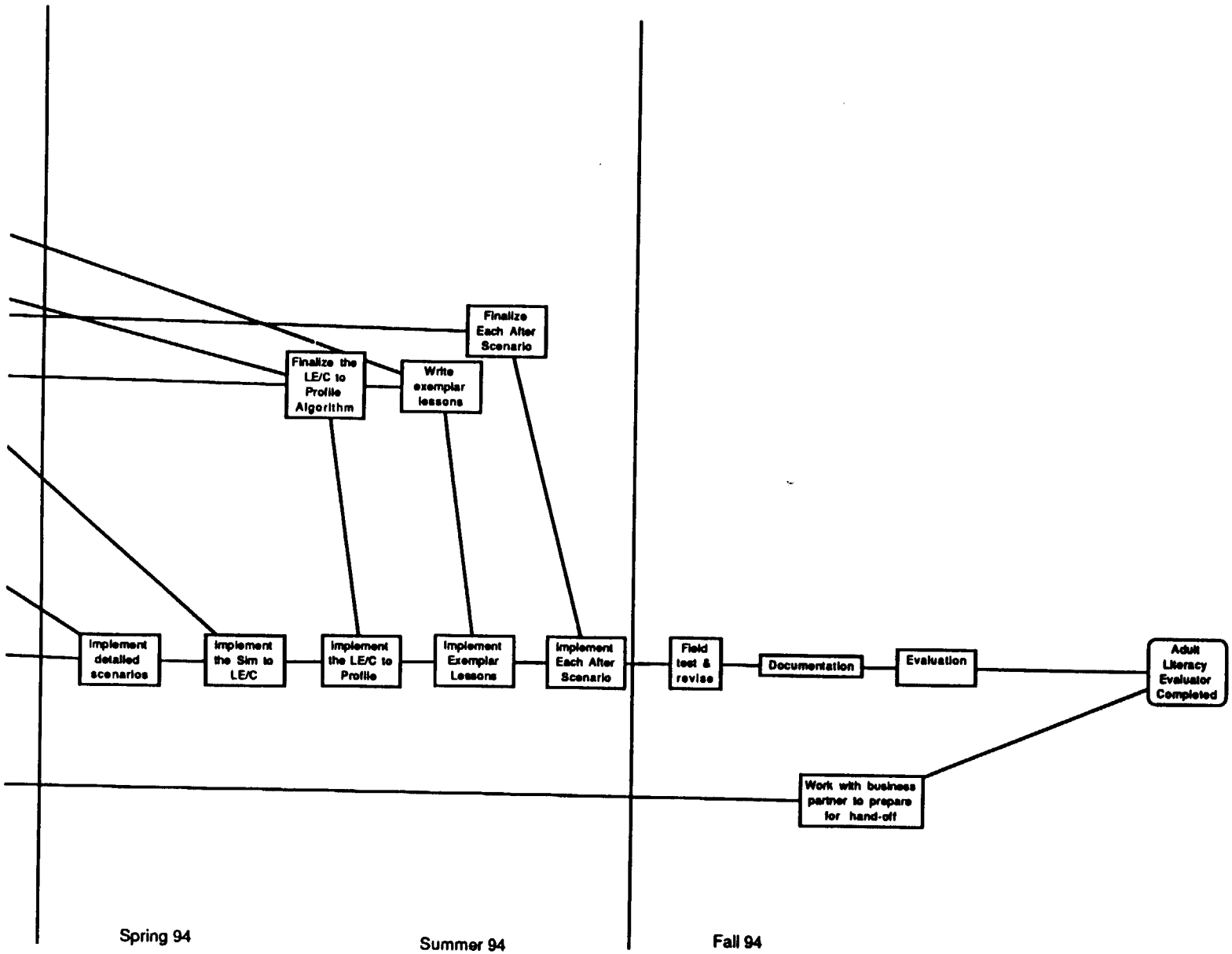
Currently, the goal of DKR is to develop a speech recognition and analysis system which will operate on a Macintosh II series computer with the first version of the system recognizing the 10 English vowel sounds in single syllable words. In this system, speech recognition is accomplished by using self-organizing feature maps (SOFM), learning vector quantization (LVQ), and hidden Markov models (HMM) operating on digitized, preprocessed speech input. The only peripheral required in this system is a MacRecorder. The first version of the system is to be delivered in October 1992.

DKR's system is best suited to the needs of ALE, partly because the developer worked on an early prototype of the system in the summer of 1991 and knows the speech analysis requirements of the system. However, the three most immediate difficulties are (a) the first version will not be delivered until October 1992, thus requiring the first prototype of ALE to be without speech recognition capabilities, (b) it will recognize vowels only and, (c) the developer cannot work full-time on the project.

Adult Literacy Evaluator Conceptual Model




Field Research
Theory & Literacy Model Development
Software Development
Grants & Quest for Business Partners
Summer 92
Fall 92 / Spring 93
Summer 93
Fall 93
Adult Literacy Evaluator
Human subject's paperwork
Interview and observe 15 pairs in field
Observe & interview clients and tutors
Tutor 3 clients bi-weekly
Form an informant group
Complete analysis of field data
Continue research on diagnostic continuums
Design an affair scenario
Preliminary dev't affair scenarios
Preliminary dev't exemplar lessons
Preliminary dev't scenarios
Preliminary algorithm dev't
Specify interrelationships of literacy elements
Specify algorithm
Finalize the LE/Cs
Finalize the Sim to LE/C Algorithm
Finalize Prelim
Design detailed scenarios
Finalize the Simulation
Implement the Simulation
Interface Experimentation
Create Field Prototype 1
Test Interface
Create Field Prototype 2
Assess and Refine Prototype 2
On-going work w/ UH development for external funding
Write grant for NLI
External funding search continues
On-going external efforts
On-going external funding
On-going search for potential business partner
Business partner quest continues
Internally search for business partner if not yet found


Finalize Each After Scenario
Finalize the LE/C to Profile Algorithm
Write exemplar lessons
Implement detailed scenarios
Implement the Sim to LE/C
Implement the LE/C to Profile
Implement Exemplar Lessons
Implement Each After Scenario
Field test & revise
Documentation
Evaluation
Adult Literacy Evaluator Completed
Work with business partner to prepare for hand-off
Spring 94
Summer 94
Fall 94

N93-26083

The Reuse of Logistics Carriers For the First Lunar Outpost Alternative Habitat Study

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1992

Johnson Space Center

Prepared By:	Carolina Vargas, undergraduate student studying Industrial Engineering.
University & Department:	Florida International University Industrial & Systems Engineering Miami, Fl. 33199
NASA/JSC	
Directorate:	Engineering
Division:	Systems Engineering
Branch:	Systems Definition
JSC Supervisor:	Kriss J. Kennedy
Date Submitted:	August 20, 1992
Contract Number:	NGT-44-005-803

The Reuse of Logistics Carriers For the First Lunar Outpost Alternative Habitat Study

ABSTRACT

The Systems Definition Branch deals with preliminary concepts/designs of various projects currently in progress at NASA. One of these projects is called the First Lunar Outpost. The First Lunar Outpost (FLO) is a proposed permanent lunar base to be located on the moon. In order to better understand the Lunar Habitat, a detailed analysis of the lunar environment as well as conceptual studies of the physical living arrangements for the support crew is necessary.

The habitat will be inhabited for a period of 45 days followed by a six month dormant period. Requirements for the habitat include radiation protection, a safe haven for occasional solar flare storms, an airlock module and consumables to support a crew of 4 with a schedule of 34 extra vehicular activities. Consumables in order to sustain a crew of four for 45 days ranges from 430 kg of food to only 15 kg for personal hygiene items. These consumables must be brought to the moon with every mission. They are transported on logistics carriers. The logistics carrier must be pressurized in order to successfully transport the consumables. Refrigeration along with other types of thermal control and variation in pressure are defined by the list of necessary consumables.

The objective of the proposed work was to collaborate the Habitat Team with their study on Logistic Carriers as possible alternatives for additional habitable volume. Options for possible reuses was also determined. From this analysis, a recommended design is proposed.

INTRODUCTION

The logistics operations for the Lunar Base are concerned with the packaging, scheduling, distribution, and storage of equipment, spares and consumables to supply the lunar surface operations.

Since the Lunar Base will be developed after the Space Station has been operational for some years, the maintenance and logistics concepts for the Lunar Base should be compatible with the Space Station system. Compatibility issues such as commonality of LRU's (Line Replaceable Unit) and dimensional standards on storage containers to assure that they can be handled and secured properly during each phase of the delivery process from the Earth to the Moon should considered. Commonality of LRU's is critical for the Lunar Base so that the same LRU will be able to support many different subsystems and subcomponents. By properly specifying common LRU's, the number of spares should be reduced and the reliability and production costs of the LRU improved.

Logistics for the Lunar Base involves three locations in a complex tracking and mass handling operation. First, the only source supplies during the first six years of lunar operations is the Earth. At a later date, oxygen supplies will be supplemented by lunar oxygen production from the pilot and production plants. The data bases which track the equipment availability, maintenance schedule, packaging, shipping and life time use will be based on the Earth and the Moon. The supplies are transported to the Space Station which serves several functions including:

- temporary holding point to prepare for final shipment to the Moon.
- minimal repackaging for shipment.
- equipment checkout and minimal refurbishment.
- source of supplies needed immediately when unavailable from the Earth.

PROBLEM STATEMENT

The purpose of the Alternative First Lunar Outpost Habitat Study is to investigate alternative habitat concepts from the base line SSF (Space Station Freedom) derived habitat. As part of this investigation, pressurized logistics modules have been identified as a possible solution to expand the habitable volume of the First Lunar Outpost.

This particular study looks at the use of logistic carriers for the purpose of expanding the pressurized volume of FLO. Included is how these logistic modules will be utilized and what functions they will support in order to use them as a habitable volume.

Some basic requirements were defined for the carrier. First of all, it is necessary to carry a resupply of consumables for a crew of four for a 45 day mission. Secondly, since 76% of the food for long duration missions is refrigerated or frozen, the carrier needs to provide refrigeration/freezer capabilities.

ASSUMPTIONS

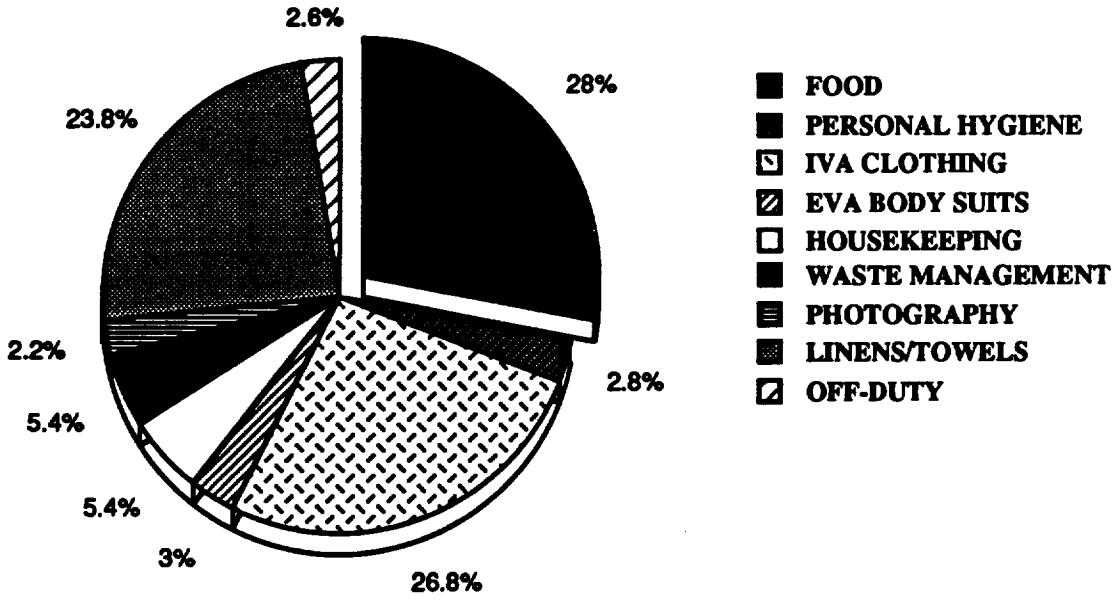
There are various assumptions that were needed to make in order to carry out the study. The habitat module is assumed to be on the ground. This assumption is necessary due to fact that the present Habitat Module is on top of the lander 50 feet above the ground. In order for the logistics carrier to be connected easily, the logistics carrier also needs to be on ground level. It is going to be transported from the landing site on a motorized undercarriage and brought to the habitat module for connection. Also, it was assumed that radiation shielding would be provided. Further analysis of radiation protection was not taken into account since it would require its own trade study.

DESIGN PROCESS

The first item that needed to be defined was the size (volume) and weight (mass) necessary in order to resupply the mission. Data was provided from the First Lunar Outpost Lunar Habitat Documentation and the Logistics Carrier Study for FLO done by Lockheed. Consumables consist of food, personal hygiene items (soap, wipes, tissues, dental care, hair grooming and body care items), IVA clothing (indoor clothing), EVA body suits, housekeeping items (dry and wet wipes, trash bags and vacuum cleaner bags), waste management (filters, fecal bags, and canisters), photography equipment, linens and towels and off-duty items (pens, pencils and other personals).

Consumables	Mass (kg)	Volume (m^3)	% Tot. vol.
Food	439	1.29	28.0 %
Personal Hygiene	52	0.13	2.8 %
IVA Clothing	190	1.24	26.8 %
EVA Body Suits	25	0.14	3.0 %
Housekeeping	79	0.25	5.4 %
Waste Management	34	0.25	5.4 %
Photography	19	0.10	2.2 %
Linens/Towels	142	1.10	23.8 %
Off- Duty	63	0.12	2.6 %
TOTALS:	**1043**	**4.62**	**100 %**

This is a chart of the breakdown of the consumables in order to determine which consumes the most space. Food is the major component while IVA Clothing and Linens/Towels rank as the second heighest consumers of module space. This is due to the fact there is no clothes washer on board and the crew members will need to use fresh items regularly.



Analysis of Design Requirements:

Design Issues:

When considering the use of the present logistic modules for possible expansion of the lunar habitat module, certain design issues must be taken into consideration. Since the habitat does not have fixed dimensions, a fair degree of flexibility concerning the design of the logistics module is adequate.

Alternate materials such as inflatables could be studied to see which material provides the optimal results and is the most advantageous. A needs analysis of pressurized vs. unpressurized requirements should be prepared in order to further define which equipment and consumables need to be under which atmospheric condition. Information on air and thermal control extensions to be connected t the proposed logistic modules could be looked at. Other minor issues such as how much trash is generated by the habitat module in order to determine the size of the waste container can add to the determining of what function will the logistic module/carrier serve.

Functions of Habitat Module:

While trying to make use of existing hardware and minimizing new technology developments, the SSF (Space Station Freedom) was chosen as the most applicable because of similar design requirements. The basic SSF module structure provides equipment that is included in the module for the LEO (lower earth orbit) application. For the lunar outpost application, some additional equipment (both internal and external to the module) is required to support the habitat and crew.

These are the different habitat segments and each of its corresponding functions:

- Communications and Data Management System (CDMS): requires extensive communications equipment capable of receiving signals which are transmitted over extremely long distances as well as the capability of sending back messages; adequate space must be allotted for the ingress of one or more crew members at a time. The CDMS is designed to provide various services such as data acquisition, data processing, data storage, subsystem control, and communications. The CDMS must support human interaction while crew is present and must be capable of operating autonomously for periods when a crew is not present (Lunar Habitat Documentation, 5/92).

- Power Distribution: provides the necessary power lines within the scarring with possible extensions for adjoining compartments.

The basic power system of the Lunar Habitat Module consists of (1) photovoltaic arrays, and (2) regenerative fuel cells. The concept is to use photovoltaic arrays for daytime energy needs and fuel cells for power during lunar nights (LHD, 5/92).

- ECLSS (Environmental Control Life Support System): recovers materials essential for crew survival, such as water or oxygen, from crew metabolic wastes such as carbon dioxide, waste hygiene water, or urine. The option selected for the lunar habitat ECLSS is the closed water loop design. This includes water processing hardware used to reclaim potable and hygiene water from crew waste water and requires only a minimum of resupply of water (91 kg) per 45 day mission (LHD, 5/92).

- Thermal Control System: provides temperature control to the habitat, inhabitants, and subsystems (internal and external to the module). this is done by systems which transport heat form the location in which it is produced to a location it can be rejected to the environment. The SSF internal Thermal Control System is applicable to the lunar habitat with a few minor changes. New external heat rejection system is required (due to the solar array temperature high of ~ 380°K during midday): a horizontal radiator with no heat pump chosen for simplicity and low power requirement (LHD, 5/92)

- EVA/Airlock: one or two independent airlock modules to support EVA operations, transfer the crew from lander to habitat, store EVA suits, as well as provide an EVA dust removal system. One of the airlocks can provide hyperbaric capabilities. Both airlocks will be able to function as safe haven for the habitat. Each airlock can accommodate two EVA crew members simultaneously during normal operations and four during contingency operations (LHD, 5/92).

- Structural Support: consists of the necessary equipment to withstand the launch loads as well as being able to hold up the Lunar Habitat once on the moon. The loads on the Lunar Habitat that result from random vibration and rotational acceleration during launching are quite large due to the greater distance of the module from the center of gravity of the Heavy Lift Launch Vehicle (HLLV). The external structures needed additionally compared with the SSF are: solar array masts, radiator supports, an antenna mast, the habitat support structure, the habitat and lander interface, and the regenerative fuel cell tank supports (LHD, 5/92).

"Living" Volumes of Habitat

- Wardroom: consists of crew gathering, eating, entertainment, and other community functions. As the focal point of the crew's environment, it serves the same purpose as a conventional living and dining room. Seating and table arrangements are flexible in order to provide possible reconfiguration according to the crew's needs. Additionally, monitors, displays, controls for both information and entertainment could be incorporated within the surrounding equipment (Initial Mars Habitat Study, 1/92).

- Galley: a food preparation facility which includes food heating/cooking equipment, short-term food storage, food preparation area, dedicated stowage for servicing equipment and utensils, clean-up facilities, handwashing facilities, drink dispenser and trash receptacles. Connections to the habitat's fresh and waste-water supply are necessary. Special task lighting will also be provided (MHS, 1/92).

- Crew Quarters: consists of private/semi-private accommodations like sleeping stations, wardrobe area, personal stowage, a shared entertainment center and work area, with a passageway to the hygiene compartment. Each sleep station can have an adjustable bunk which can be reconfigured to allow a sit-up position. Additional stowage either above or below the sleeping stations is also a consideration (IHMS, 1/92).

- Hygiene Facility: includes two habitat facilities integrated into a common area - personal hygiene and body waste management. Although both functions are addressed within the same enclosed volume, each should be independent subsystems which could be accessed simultaneously in private. The hygiene facility will include provisions for whole and partial body cleansing, handwashing, and personal grooming (IMHS, 1/92).

- Exercise/Health Facility: dedicated region within the module which serves both routine crew exercise and health care needs as well as emergency medical situations. In emergency situations, certain surrounding portions could be removed in order to provide easier ingress. When health care treatment is not required, practically the entire region could accommodate exercise activity. The equipment and electronics volumes could be located overhead to maximize volume (IMHS, 1/92).

- Trash Management Facility: it incorporates trash collection and compaction functions while providing an easy transfer of the compacted trash to a more permanent storage location. Other receptacles for trash collection may be located in the crew quarters as well as the hygiene facility for intermediate collection

until transferred to this facility for compaction. Perhaps some sort of energy source could be derived from the elimination of this trash (IMHS, 1/92).

» Food/Consumables Stowage: adjacent to the galley and food preparation area are subsystems of palletized food packages further divided into individual portions of food items in long-term stowage. Certain food items may be transferred in a short-term containment facility or refrigerator within the galley for more convenient access (IMHS, 1/92).

» Laundry Facility: cleaning of personal clothing, bedding, and towels will be performed here. The needed equipment includes a washer/dryer, intermediate stowage of launderable items, consumables used in the process, and a suitable surface for organization and folding of the laundry. Collection points within the habitat will transfer to this facility for laundering (IMHS, 1/92).

» Additional Stowage: provides extra room for equipment, supplies, personal items, and/or cleaning utensils. Goods from the logistics carriers could be temporarily stored here.

From all of these habitable volumes, it is clearer which functions could be retrofitted for the Logistics Carrier. Yet, certain sizing dimensions still need to be addressed.

SPACE STATION DATA

Boeing Aerospace Company provided Marshall Space Flight Center with a document which defines the known configuration of the Logistics System required to support the Space Station program. This report provides both mass and volume requirements necessary for a resupply to Space Station. From this data, an analysis of the ratios of its components to the total volume allows the determination of the total necessary volume needed for FLO.

This is the overall breakdown:

Total Internal Volume:	**85 m^3**	**% of Total**
Consumables	43.5	51.2
Access Space	30.5	35.9
Subsystems	11.0	12.9

CALCULATION OF TOTAL VOLUME

These ratios provided a basis upon which to calculate the volume necessary for the determined volume of consumables. Each value determines the approximate volume taken up by each category. This information was necessary in order to determine the total volume for both resupply and habitation purposes.

As stated previously, the total consumables required for a crew of four for a 45 day mission is **4.62 m^3**. Adding a factor of 20% for a packaging void, an additional **0.924 m^3** is added to the current amount in order to allow for unaccounted space. The total comes out to **5.54 m^3**. This total is then divided by the ratio of consumables for the Space Station module. The volume for the consumables, **5.54 m^3** divided by 51.2 % (SSF's consumables ratio) gives a grand total of **10.8 m^3**. From the total, the following breakdown was developed.

Volume Requirements:

Consumables (+ 20% packaging void)	**5.54 m^3**
Subsystems Volume	**1.40 m^3**
Access Space	**3.86 m^3**

SHAPE STUDY

Of the shapes to be considered in order to decide on the geometry of the carrier, three main shapes were identified: a sphere, a rectangle, and a cylinder. The sphere did not work well with any rack system since consumables usually took on some sort of rectangular shape. The rectanglular shape does not work well with a pressurized environment. A cylinder, either vertical or horizontal provided the most positive feedback.

In considering the feasible solutions between a horizontal or vertical cylinder, various items were considered. The volume space of both configurations differs in usability. The horizontal orientation provides 5.5 m^3 rack space for the consumables. While the vertical orientation only provides 3.4 m^3 of available rack space due to the gaps of unaccessible space. Circulation space in the horizontal mode works better for reconfiguration than the vertical mode.

DESIGN REQUIREMENTS

Once the horizontal configuration was chosen to be the determined shape, a few basic design requirements were needed to be made. The height was determined based on data from Skylab. Their vehicle was six and a half feet high in zero gravity. Allowing for some additional space due to the 1/6th g, a height of seven feet six inches was decided to be adequate. The access width was defined by physical dimensions of a man in an EVA suit. The maximum breadth for a 95%ile male is 34 inches. That is the exact width of the proposed design. Since the carrier needs to be pressurized in order carry the consumables, the shape must be in accordance. Lastly, when the logistics carrier is no longer self-sufficient, it must have the capability to be dependent on the habitat. For figures of proposed design see appendix.

ADDITIONAL REUSES

Possibilities for reuse were considered only of the functions which would not cause too much of an additional load on the habitat's present systems. Crew quarters is a strong possibility due to the fact that little additional power was needed to reconfigure the carrier. Science laboratories is also an option since there is refrigeration and freezer capabilities already on board. EVA storage and maintenance could be applied in the carrier and leave that much more extra space on the habitat module. Last but not least, it could be used as equipment stowage.

CONCLUSIONS

From this study, certain objectives were attained. Information on the details of the logistics operations was provided to the Alternative Habitat Team. The insight on possible reuses helps to determine the real possibilities of the logistics carriers reusability. Since a logistics carrier is needed regardless if it is reused or not, it would be more volume efficient to use its space once its original functions have been accomplished. Also, added space helps to improve the psychological and social attitudes between each crew member. The flexible configurations such as being able to remove the racks enhances the suitability for each mission. As a long term goal, it opens the door for longer duration missions with the possibility of permanent residence on the moon.

REFERENCES

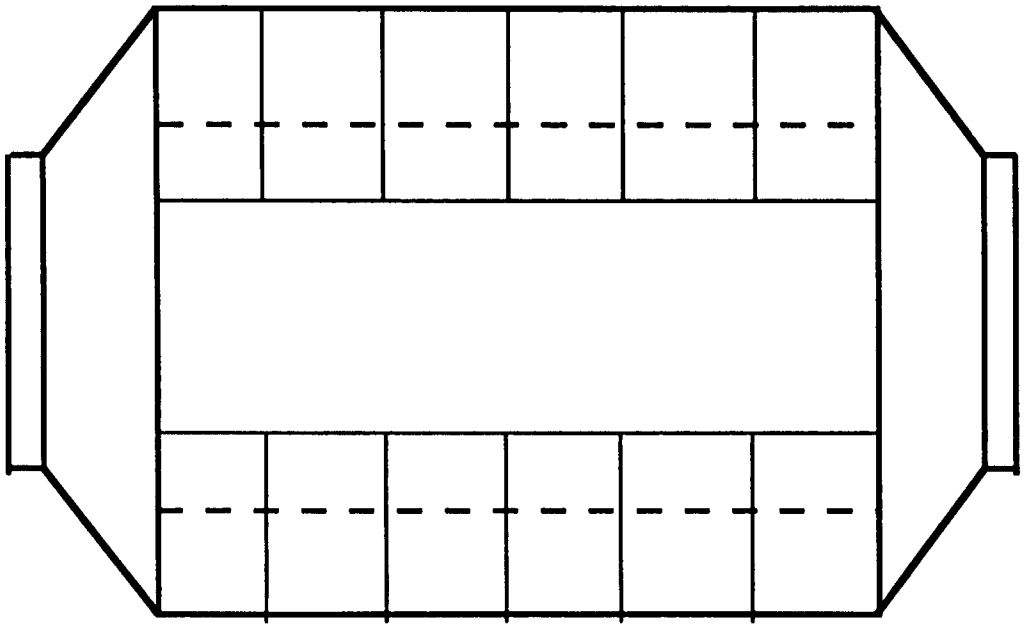
Logistics Module End Item Data Book. Boeing Aerospace Company Space Station Program. Contract no. NAS8-36526

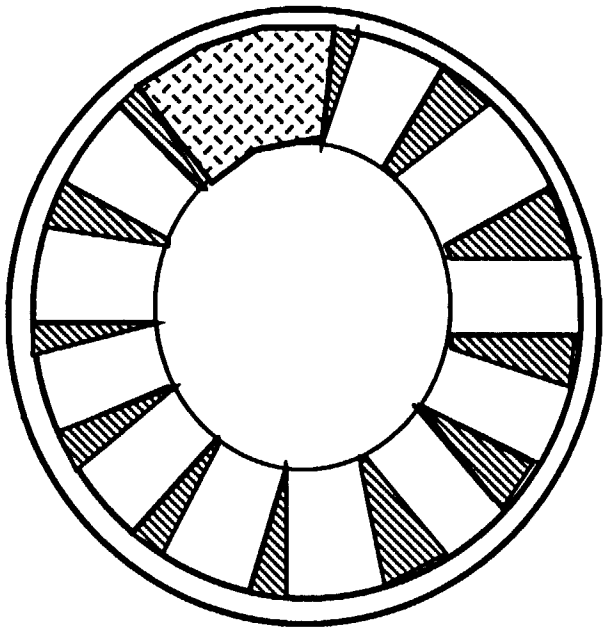
Lunar Habitat Documentation. Marshall Space Flight Center. Huntsville, Alabama. May 1992.

Lunar Base Maintenance and Supply Options. Advanced Space Transportation Support Contract. No. NAS9-17878.

Proposed Concepts

Horizontal **vs.** **Vertical**

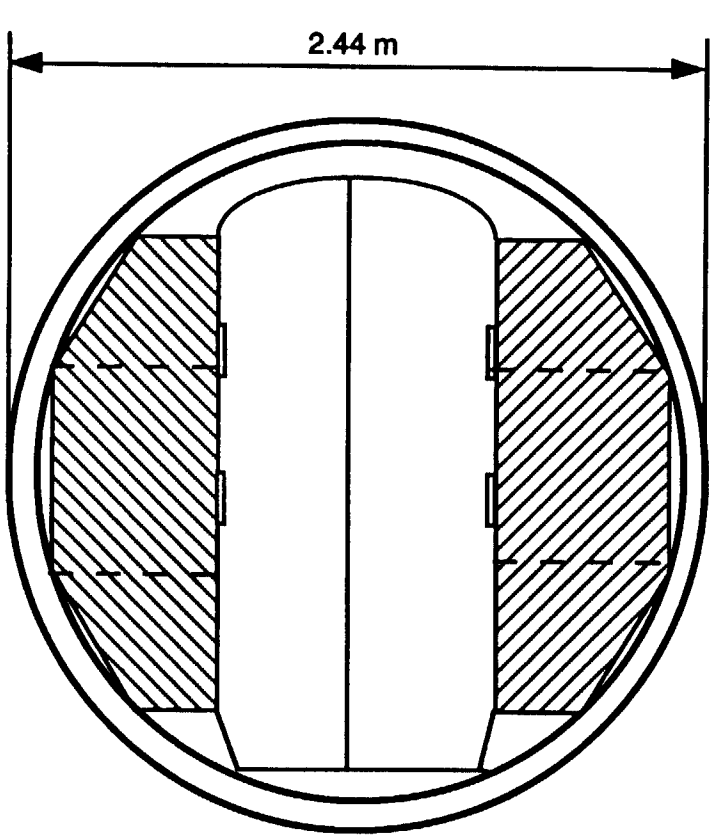




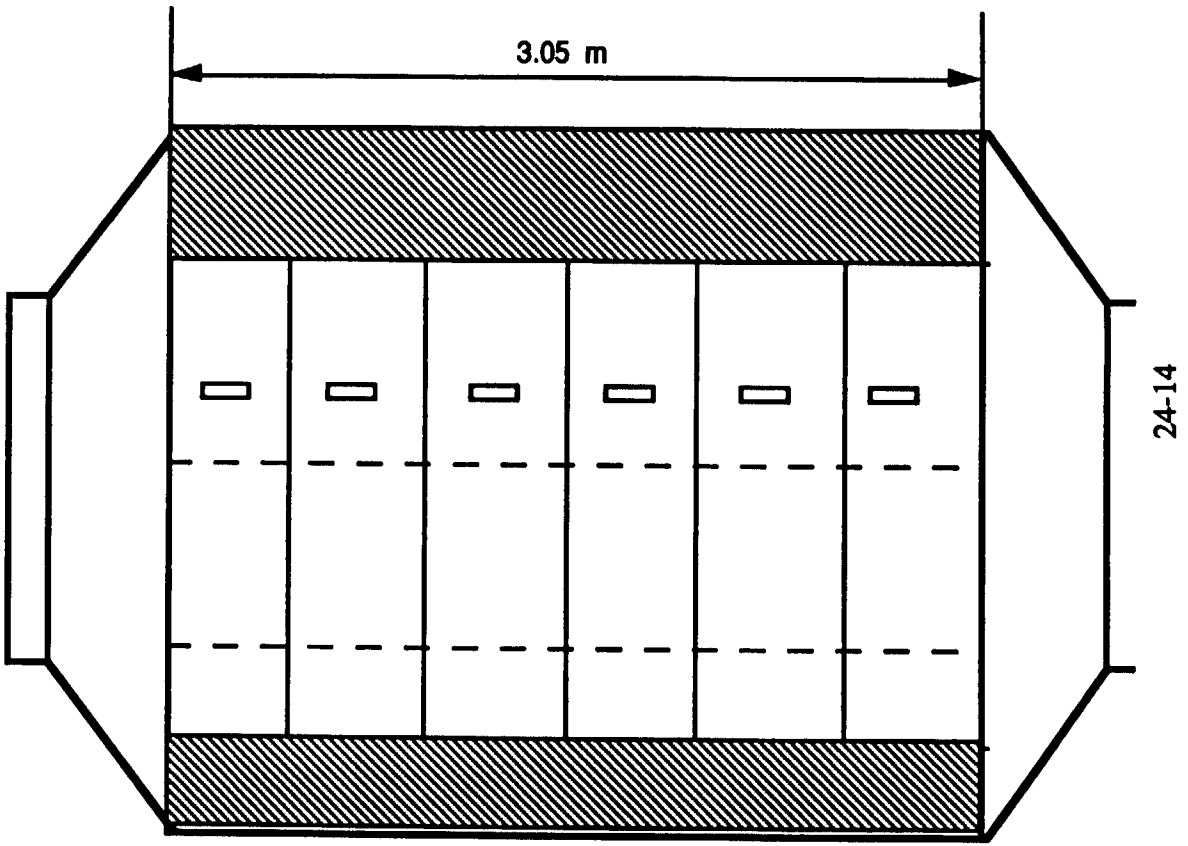
PLAN VIEW

24-13

Recommended Design



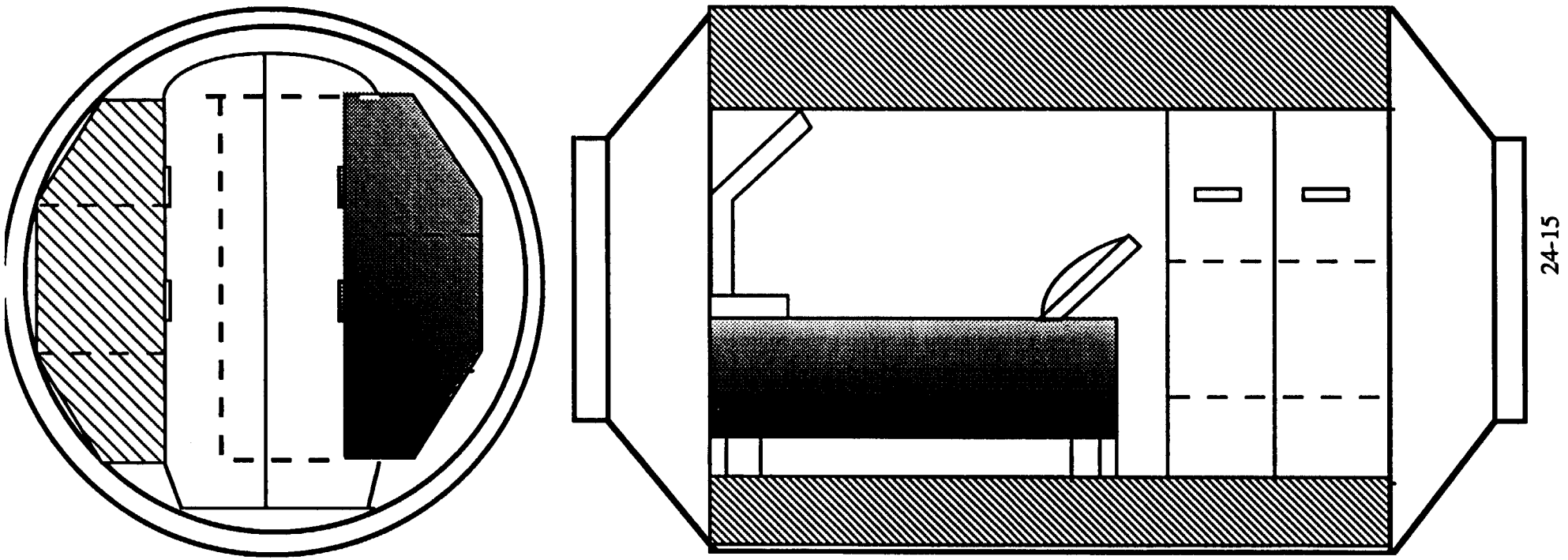
Front View



Side View

Reuse Options

- Racks removed to provide crew quarters



24-15

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1992	Contractor Report

4. TITLE AND SUBTITLE
National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program–1992 Volume 2

5. FUNDING NUMBERS
NGT 44-005-803

6. AUTHOR(S)
Richard B. Bannerot and Stanley H. Goldstein, Editors

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Houston
Houston, Texas

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
University Programs Office
Lyndon B. Johnson Space Center
Houston, Texas

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
NASA CR 188242

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Unclassified/Unlimited
Star Category 99

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The 1992 Johnson Space Center (JSC) National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program was conducted by the University of Houston and JSC. The program at JSC, as well as the programs at other NASA Centers, was funded by the Office of University Affairs, NASA Headquarters Washington, D.C. The objectives of the program, which began nationally in 1964 and at JSC in 1965, are (1) to further the professional knowledge of qualified engineering and science faculty members; (2) to stimulate an exchange of ideas between participants and NASA; (3) to enrich and refresh the research and teaching activities of participants' institutions; and (4) to contribute to the research objective of the NASA Centers.

Each faculty fellow spent at least 10 weeks at JSC engaged in a research project in collaboration with a NASA JSC colleague. This document is a compilation of the final reports on the research projects performed by the faculty fellows during the summer of 1992. Volume 1 contains reports 1 through 12, and Volume 2 contains reports 13 through 24.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19 SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102